

Durham Research Online

Deposited in DRO:

03 November 2020

Version of attached file:

Published Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Cartwright, N. (2020) 'Lullius Lectures 2018 : Mid-level theory : without it what could anyone do?', *Theoria*., 35 (3). pp. 269-323.

Further information on publisher's website:

<https://doi.org/10.1387/theoria.21479>

Publisher's copyright statement:

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License.

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.



Middle-range theory: Without it what could anyone do?

(*Teoría del rango medio: ¿qué haríamos sin ella?*)

Nancy CARTWRIGHT*

Durham University and University of California at San Diego

ABSTRACT: Philosophers of science have had little to say about ‘middle-range theory’ although much of what is done in science and of what drives its successes falls under that label. These lectures aim to spark an interest in the topic and to lay groundwork for further research on it. ‘Middle’ in ‘middle range’ is with respect to the level both of abstraction and generality. Much middle-range theory is about things that come under the label ‘mechanism’. The lectures explore three different kinds of mechanism: *structural mechanisms* or *underlying systems* that afford causal pathways; *causal-chain mechanisms* that are represented in what in policy contexts are called ‘theories of change’ and for which I give an extended account following the causal process theory of Wesley Salmon; and *middle-range-law mechanisms* like those discussed by Jon Elster, which I claim are—and rightly are—rampant throughout the social sciences. The theory of the democratic peace, that democracies do not go to war with democracies, serves as a running example. The discussions build up to the start of, first, an argument that reliability in social (and natural) science depends not so much on evidence as it does on the support of a *virtuous tangle of practices* (without which there couldn’t even be evidence), and second, a defence of a *community-practice centred instrumentalist* understanding of many of the central basic principles that we use (often successfully) in social (and in natural) science for explanation, prediction and evaluation.

KEYWORDS: evidence-based policy (EBP), middle-range theory, mechanism, marker, theory of change, at-at theory, democratic peace theory, support factors, realist evaluation, Jon Elster, Wesley Salmon, community-practice centred instrumentalism, argument theory of evidence.

RESUMEN: Los filósofos de la ciencia han tenido poco que decir acerca de la “teoría de rango medio”, aunque gran parte de lo que se hace en la ciencia y de lo que impulsa sus éxitos cae bajo esa etiqueta. Estas conferencias tienen como objetivo despertar el interés en el tema y sentar las bases para la ulterior investigación al respecto. “Medio” en “rango medio” hace referencia al nivel de abstracción y generalidad. Gran parte de la teoría de rango medio trata sobre cosas que caen bajo la etiqueta de “mecanismo”. Las conferencias exploran tres tipos diferentes de mecanismos: mecanismos estructurales o sistemas subyacentes que permiten vías causales; mecanismos de cadena causal que están representados en lo que en contextos de política se denominan “teorías de cambio” y sobre los cuales doy una explicación extensa siguiendo la teoría del proceso causal de Wesley Salmon; y mecanismos de ley de rango medio como los discutidos por Jon Elster, que afirmo son, y con razón son, rampantes en todas las ciencias sociales. La teoría de la paz democrática, que las democracias no van a la guerra con las democracias, sirve como un ejemplo en funcionamiento. Las discusiones se desarrollan hasta el comienzo de, primero, un argumento según el que la confiabilidad en las ciencias sociales (y naturales) no depende tanto de la evidencia como del apoyo de una maraña virtuosa de prácticas (sin las cuales ni siquiera podría haber evidencia), y en segundo lugar, una defensa de una comprensión instrumentalista, centrada en la práctica comunitaria, de muchos de los principios básicos centrales que usamos (a menudo con éxito) en ciencias sociales (y naturales) para explicar, predecir y evaluar.

PALABRAS CLAVE: política basada en la evidencia (PBE), teoría del rango medio, mecanismo, marcador, teoría del cambio, teoría at-at, teoría de la paz democrática, factores de apoyo, evaluación realista, Jon Elster, Wesley Salmon, instrumentalismo centrado en la práctica comunitaria, teoría de la argumentación de la evidencia.

* **Correspondence to:** Nancy Cartwright. Department of Philosophy, Durham University, 50 Old Elvet, DH1 3HN (United Kingdom) – nancy.cartwright@durham.ac.uk – https://orcid.org/0000-0002-0873-6966

How to cite: Cartwright, Nancy (2020). «Middle-range theory: Without it what could anyone do?»; *Theoria. An International Journal for Theory, History and Foundations of Science*, 35(3), 269-323. (https://doi.org/10.1387/theoria.21479).

Received: 18 February, 2020; Final version: 12 March, 2020.

ISSN 0495-4548 – eISSN 2171-679X / © 2020 UPV/EHU



This article is distributed under the terms of the
Creative Commons Attribution 4.0 International License

Lecture 1. Markers, mechanisms & measures

MIDDLE-RANGE THEORIES —THE ONLY GAME IN TOWN

These lectures aim to get you excited about *middle-range theories*. We have little philosophical understanding of what they are, what they do or of how they do what they do. This despite the fact that middle-range theorising is ubiquitous in the social, economic and health sciences.¹ There's good practical reason why it is so widespread: if we are concerned with explaining, predicting or managing the world in these domains, middle-range theories are generally the only game in town.

Middle-range theory is everywhere in the social sciences, including economics, which is notorious for its use of abstract, mathematised general theories. Philosopher/historian Mary Morgan provides a nice example of this in her paper 'The Curious Case of the Prisoner's Dilemma', where she notes, 'Game thinking has introduced a form of reasoning into economics that adds case-based reasoning onto more generalizing kinds of theory' (Morgan, 2007, p. 61). Game theory models are examples of middle-range theorising par excellence.

What though is 'middle-range theory'? The term was made popular by the sociologist Robert Merton in the 1940s. By now it means many different things to many different people. We will get a better sense of some of the important scientific endeavours that fall under this label as these lectures proceed, but for a start you can think of them as the sociologists Peter Hedström and Lars Udehn describe. Middle-range theory is

... a clear, precise, and simple type of theory which can be used for partially explaining a range of different phenomena, but which makes no pretense of being able to explain all social phenomena... It is a vision of sociological theory as a toolbox of semigeneral theories each of which is adequate for explaining a limited range or type of phenomena. ...[Merton's] vision has more in common with the type of theories found in the life sciences than with those found in the physical sciences. Contemporary biologists, for example, do not seek to develop general law-like theory but aim at identifying middle-range theory in Merton's sense of the term. (Bearman and Hedström, 2011, p. 31)

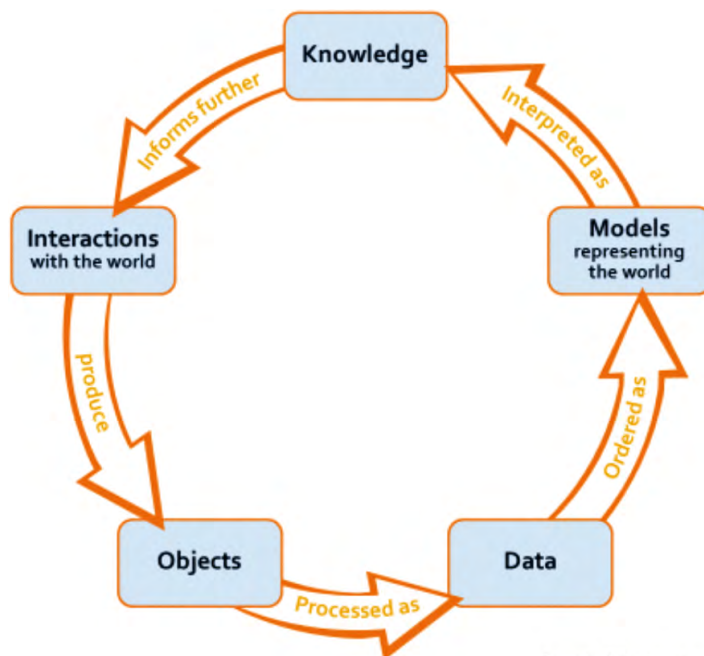
A lot of middle-range theory goes under the label 'mechanism'. This will be one of the central themes of these lectures. 'Mechanism' is a term that is used in a great variety of ways both in philosophy and across the sciences. I am going to sort out three distinct kinds of mechanism that play a significant role in middle-range theorising in social science. First are *structural mechanisms*, which I shall discuss here in Lecture 1. This is a sense of 'mechanism' familiar to philosophers from work in the last two decades in the philosophy of biology. Second are *causal-chain mechanisms*. Starting with the process/interaction theory of Wesley Salmon Lecture 2 will develop an account of what a good mechanistic theory of this kind should be like. The third, which I turn to in Lecture 3, are *middle-range-law mechanisms*. These may be familiar to philosophers from the work of Jon Elster (2015, p. 56), who talks of social and psychological mechanisms, which often come in 'opposing pairs, like the *endowment effect*, in which "a memory of a good experience is a good mem-

¹ It is also, in my view, in physics as well. But I will not defend that here.

ory” and the “experience of a bad [memory] is a bad memory” suggests opposite outcomes to those of the *contrast effect*, in which memories of past experiences have an opposing effect on present experiences.’

I am going to use as a running example of a middle-range theory the theory of the democratic peace (*DPT*) —that democracies do not go to war with democracies. Democratic peace theorising is a rich, variegated activity with a great many interlocking endeavours of various kinds, some of which I shall review.

Middle-range theorising comes in a variety of forms, involving a variety of practices serving a variety of purposes. I can illustrate using some slides lent to me by Sabina Leonelli from her 2018 Lakatos Award lecture at the London School of Economics. The first (Figure 1) shows a cycle of activities that go into knowledge creation in the sciences. Leonelli’s second slide (Figure 2) points out that *theorising* is required for all of them. And, I add, by far and away the bulk of this theorising will be middle range. These lectures will focus on just a few of these. Today, in Lecture 1 I am going first to introduce you to *markers* that middle-range theories like the democratic peace can provide for us. Markers are characteristics, like *democracy*, by which we identify systems that support or, alternatively, discourage causal pathways we are concerned about, like the route from tension to war. Second, I am going to provide a philosophic framework for one kind of middle-range theorising that I have looked at in some detail: the construction and defence of measures for theoretical concepts.



Leonelli S (forthcoming) What Distinguishes Data from Models?
European Journal for the Philosophy of Science

Figure 1

Leonelli (2019, p. 22)

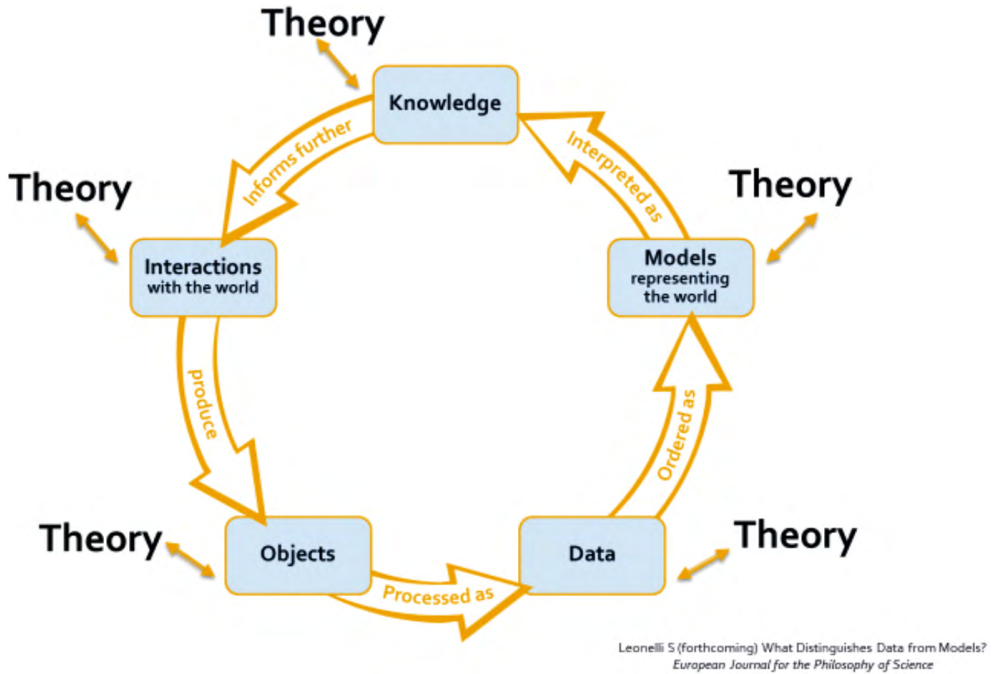


Figure 2

Leonelli (2019, p. 22)

The topic of Lecture 2 is *theories of change* and *causal-chain models*. Lecture 3 turns to mechanisms in the *middle-range law* sense. These are rife in the social sciences. Yet they seem far too loose to make for good science, especially since, as Elster notes, many point in opposite directions. Elster himself says that, though they are useful for explanation, they will not serve for prediction. But if they are so loose that we cannot use them for prediction, why suppose they can participate in credible explanation? I shall argue that explanations employing Elster-type middle-range laws can be credible —if they are embedded in a thick, tangled net of middle-range practices like that we find in democratic peace theorising.

WHAT WORKS: MAKING SURE YOUR POLICIES DELIVER WHAT YOU WANT

Much of my recent work has been in the philosophy of social technology: on how to use scientific and local knowledge to build better societies, with special focus on better policies and better predictions about policy outcomes. This is how I came to middle-range theory. I'll start there with you too.

I began this work because I was disturbed by the evidence-based-policy (EBP) movement, which is very vocal in the US and the UK as well as in various international organisations (like the World Bank's Open Knowledge Repository). For instance, the UK now has

nine expensive ‘What Works’ Centres, and in the US what counts as “scientifically-based research in education” was written into Federal law in the ‘Castle’ Bill of 2002,² which sets the standards for the US Department of Education’s What Works Clearinghouse. I was disturbed on two counts.

First is the very narrow range of scientific work these EBP sites take on board. Most set themselves the task of vetting, summarising and disseminating the ‘evidence’ about how good the policies or ‘interventions’ they review are at producing their targeted outcomes. EBP institutions do so in order to help policy makers make better decisions about what policies to implement. But their notion of ‘evidence’ is highly restrictive. It does not cross the range of credible scientific knowledge. Rather, almost everywhere EBP has concerned itself with policing the quality of *empirical studies*, and a very narrow kind of empirical study at that. The focus is almost entirely on studies designed to look directly at whether specified *inputs* produce targeted *outputs* in study populations, and within those, on studies that make statistical group comparisons, whether randomised controlled trials (RCTs), which are the gold standard across EBP, or observational studies comparing groups where the input occurred versus ones where it did not.

This restriction on the scientific knowledge base and on the kind of scientific endeavour that is licensed is generally defended on the grounds that *a causal claim must be established by establishing a counterfactual*. To this is added the claim that it is impossible to establish a counterfactual for a single individual since we cannot look to see both what happens if that individual is subject to the input and if not. But in a good RCT design we can at least get an unbiased estimate of what happens in two groups where the input is independent of other causal factors, and that’s what they advocate. But note: this is one big step removed from establishing a well-formed counterfactual. At best³ what an RCT warrants is *that* a counterfactual of the right form holds for some members of the study population; it does not establish any specific counterfactual. So at best the conclusion can be: for somebody in the study population, the specified input produced the targeted output.

We philosophers have had a lot to say, pro and con, about causes and counterfactuals. What I want to note here is that, even if we assume that causal claims amount to counterfactual claims,⁴ it does not follow that causal/counterfactual claims must be established by such a direct line of argument. Depending on what we assume about causality and what we take to be known about the type of situation at hand, there are a large variety of ways of establishing a causal claim and thereby —if one assumes the equation of causes with counterfactuals— a counterfactual. But ‘what we take to be known’ will involve a great deal more than study results of the kind EBP majors in. In particular, it will involve ‘middle-range’ theory.

My second shock in engaging with EBP was with the exceedingly sloppy language employed throughout. According to conventional EBP wisdom, by careful review of the research evidence we can learn “What Works” and (setting aside questions of cost and cultural and moral acceptability) that is what policy makers should implement.

² For an account of the federal legislation here see Eisenhart and Towne (2003).

³ ‘At best’ because the RCT result is only an (unbiased) *estimate* of the true average treatment effect. If the true average is greater than zero, it follows that the individual treatment effect, which is defined in terms of the individual counterfactual, is positive for at least some individual studied.

⁴ Which I do not endorse.

But: what on earth could “It works” mean? I will here pass over our usual thorny issues about causation in general in order to focus on further questions of serious practical consequences. “It works” sounds like a rubric for a general claim: “This policy results in the targeted outcomes generally” or perhaps “... widely” or “...across this domain”, where what ‘this’ domain is may be more or less clear. That can’t, though, in good faith be right because of the way these claims are evidenced. They are taken to be well established if one or a number of good group comparison studies all show that the policy achieved its targeted outcomes in the study populations in the study sites and none show they did not. And this without seriously attending to any of our usual worries in philosophy of science about how to establish general claims, like cautions about induction by simple enumeration (it worked in site 1, it worked in site 2, ... so it will work everywhere), the need to ensure we are using projectable predicates and so forth.⁵

Equally under warranted by evidence that the policy has worked in some number of study sites is the usual advice given to policy makers: this policy works, so use it. But of course everyone recognises that the fact that a policy worked in some places is poor guarantee that it will work in this new place: “Context matters.” This has now become a familiar slogan in the EBP community. We should, we are now told, not just ask “What works?” but “What works, for whom, where.” The problem is that the science condoned in EBP is not up to the job of answering that.

But middle-range theory can help. That’s why I became interested in it. In these lectures I shall investigate various types of scientific work reasonably labelled ‘middle-range’ theory that can support causal prediction, exploring how they work and some of the problems in their use. Today I start at the deep end— literally, with social substructures, because that is what in the end determines what can cause what.

CONTEXT MATTERS

‘What Works, for whom, where?’ Answers to this come in the form of conditional general claims, although not always explicitly so formulated. Here are three typical cases that I have studied.

- ‘Rigorous evidence shows that *school-based deworming can improve children’s health, education, and long-term productivity at an average cost of less than \$0.50 per child per year*’ (italics added).⁶ This is from the website for *Evidence Action*, an anti-poverty institution that says of itself: ‘By focusing on promising approaches backed by rigorous evidence, we support programs that are measurably effective.’ You may have heard of this kind of institution from the ‘Give effectively’ movement.

The italicised phrase sounds like a general claim, one that is presumably meant to hold conditional on at least the fact that the children are in areas affected by parasitic worms since

⁵ For a discussion of this problem see Cartwright and Hardie (Oxford University Press, 2012) on ‘climbing up the ladder of abstraction’, especially the example of the Bangladesh and Tamil Nadu Integrated Nutrition Programmes.

⁶ The Problem: Parasitic Worms. *Evidence Action*. Retrieved from <https://www.evidenceaction.org/dewormtheworld/#the-problem-of-parasitic-worms>

they tell us: ‘While virtually non-existent and unheard of in developed countries today, parasitic worm infections are endemic in many of the poorest countries in the world.’

- With respect to the effects of the multi-faceted literacy programme Read 180 on comprehension: there is ‘Strong evidence of a positive effect with no overriding contrary evidence.’⁷ This is according to The US Department of Education’s Institute of Education Sciences’ *What Works Clearing House* (WWC), which claims of itself: ‘We focus on the results from high-quality research to answer the question “What works in education?”’⁸ The WWC effectiveness ratings, like the one for *Read 180*, are based on ‘the quality of research, the statistical significance of findings, the magnitude of findings, and the consistency of findings across studies’.⁹
- ‘The Signs of Safety approach provides principles, disciplines and fit-for-purpose tools that equip practitioners and supervisors to build observable everyday safety for children...’ This, in the realm of child protection, is from the *Signs of Safety Comprehensive Briefing Paper* (Turnell, 2019), which explains what Signs of Safety is and how and why it should work. This approach is being adopted in many local areas in the UK and the Republic of Ireland, following the 2011 *Munro Review of Child Protection*, which identified interventions into the child welfare system that did not work to produce better outcomes for children: ‘statutory guidance, targets and local rules’, which, the *Review* maintains ‘have become so extensive that they limit [the] ability [of practitioners] to stay child-centred. The demands of bureaucracy have reduced their capacity to work directly with children, young people and families... Services have become so standardised that they do not provide the required range of responses to the variety of need that is presented’ (Munro, 2011).

Whatever one thinks about basic laws of nature, general claims like these about what policies produce what outcomes are not among them. That’s why context is so critical —the *where* part of ‘What works, for whom, where’ really matters. *Ceteris paribus* claims like these hold some places and not others, and there is a reason for that. They hold because there is some underlying structure (sometimes also referred to by the word ‘system’) that gives rise to them. They are afforded by the basic materiel, social, economic and cultural arrangements that obtain, and these differ from place to place.

I have called these underlying arrangements ‘socio-economic machines’ or, more generally, ‘nomological machines’, because they generate law-like (*ceteris paribus*) regularities. Socio-economic machines are akin, in the social realm, to the ‘mechanisms’ that philosophers of biology like Peter Machamer, Lindley Darden and Carl Craver (MDC), William Bechtel, Stuart Glennan and others cite to explain biological regularities. As Rom Harré argued already in 1972 different underlying structures afford different causal pathways and in consequence, different causal regularities (Harré, 1972). I shall call these kinds of mechanisms ‘structural mechanisms’ to distinguish them from other kinds I will discuss in these lectures. My standard household example is the toilet and the toaster. Pressing the lever has

⁷ WWC Intervention Report. *READ 180*, November 2016. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_read180_112916.pdf

⁸ What Works Clearing House. Retrieved from <https://ies.ed.gov/ncee/wwc/>

⁹ WWC Summary of Evidence for this Intervention. *READ 180*. Retrieved from <https://ies.ed.gov/ncee/wwc/intervention/742>. Accessed 30 March 2019.

very different outcomes in the two, and that depends on the parts of the underlying structure, their arrangements and activities.

The same dependence on underlying mechanisms is true for singular causal relations as well as for *ceteris paribus* causal regularities. For most singular causal facts of interest to us there is seldom anything in the cause directing it to that effect. This is especially true when the cause is a policy intervention and the effect is its targeted outcome, as in my three examples earlier. In these cases there are long chains of intervening steps. It is the arrangements and activities of the parts of the structure in which the cause operates that allows

- that each step can produce the one that follows and
- the concatenation of the steps to allow the entire process to carry through.

Lecture 2 will discuss these causal chains in some detail. I mention them here to underline the idea that, when it comes to claims of practical use in managing the world around us, whether we consider a *ceteris paribus* causal regularity claim or a singular causal prediction about what will cause what in a new case (or a retrodiction about what has done so), it is the structural features of the context that determine the truth of the claim.

This makes for real problems, for we have no explicit methodology for how to study these underlying mechanisms in order to determine what causal pathways they afford. This is not to say that we don't study them. We do, and sometimes we do a very good job of it—as the studies discussed by the mechanist philosophers of biology witness. Rather, there aren't methods handbooks on it,¹⁰ nor are there good accounts of appropriate methodologies in our own discipline of philosophy of science. What I have come to see in investigating these problems is that a good deal of middle-range theorising of various sorts is involved in our successful endeavours at this. In particular when it comes to the special case of policy prediction, middle-range theory can help in two ways:

- It can identify what I label 'markers' that pick out structures that will support a given causal pathway.
- It can tackle directly the problem of identifying what the central features are of those structures that allow them to do so.

These are both topics I shall take up, beginning with the first today and turning to the second in Lecture 3, where I discuss Elster-style 'middle-range-law' mechanisms.

In earlier work I have emphasised the importance of markers in helping us negotiate the world. *Markers* are easily recognisable signs that allow us to pick out systems that have the right structure to support causal pathways of interest without our having to understand why—what features and activities allow those systems to do so, as for instance in the visible differences between acorns and nasturtium seeds. I will illustrate with the case of democratic peace theory. I pick this example because Sharon Crasnow (2012) has made it familiar in the philosophic literature through her work on case studies and because colleagues in international development with whom I have discussed markers have found the democratic peace to be a helpful example.¹¹

¹⁰ Though methods courses on case studies and causal process tracing do bits of the job.

¹¹ See Cartwright *et al.* (2019).

MARKERS AND THE CASE OF THE DEMOCRATIC PEACE

Democratic peace theory revolves around the middle-range generalization that democracies do not go to war with democracies. There is widespread agreement that the empirical claim is true but considerably less agreement on why, though there are a number of explanations on offer, which we shall look at later. To the extent that this claim is credible, it can be useful in explanation, and supposing the democracy mechanism is the principal one that will operate, at predicting and planning future actions, for instance in military policy or international investment policy.

The claim ‘democracies do not go to war with democracies’ is an example of a middle-range theory par excellence. It is in the middle with respect to the abstractness of the concepts employed and the breadth of the claim’s applicability, i.e. between high social theory and more specific social science claims that use more detailed concepts about more specifically identified issues. On one side stand broad perspectives: structuralist perspectives (like functionalism and Marxism), interpretivist ones (like interactionism), feminist perspectives, New Right perspectives or post modernism, with below these more detailed but still ‘grand’ theories like those of Antonio Gramsci, Tony Giddens, Theda Skocpol, Pierre Bourdieu, Emil Durkheim, Judith Butler and so forth. On the other side stand more specific, descriptive claims, like this typical remark about the Fashoda crisis, an incident in 1898 that is often cited as a paradigmatic example of the democratic peace in action: ‘[T]he French fleet was out-powered by superior British naval prowess and France was eager to have Britain’s support in future conflicts against Germany. The reopening of the Dreyfus Affair also distracted the French from the Fashoda dispute. Seeing no advantage in a colonial war with Britain, the French ultimately decided to withdraw its soldiers from Fashoda...’¹²

If we are to use ‘democracy pair’ as a policy marker for countries in which tensions do not result in war, we better have good reason to think that this application of DPT is accurate, or — accurate enough for this purpose. This is where middle-range theorising comes into its own. And elements of the theory itself cast some doubt on this. One school of thought accepts that democracies do not go to war with democracies but urges that that is not due to the nature of democracy. Democracy is just co-extensional with the genuine causes of dyadic peace. For instance, Henry Faber and Joanne Gowa (1995) argue that the Cold War largely explains the democratic peace, and Raymond Cohen (1994) argues that it is shared culture not democracy that matters, whereas Bruce Russett is at pains to show the contrary, that the democratic peace is ‘a result of some features of a democracy, rather than caused exclusively by economic or geopolitical characteristics correlated with democracy’ (1994, p. 11), like alliances, political stability or wealth.

This matters to the usefulness of the democracy marker for policy prediction and planning. For instance, Russett, like many US Presidents (Owen IV, 2005), suggests that US efforts to democratise other countries may not be just for democracy for its own sake but also to make the world more peaceful. This is in agreement with Sebastian Rosato, who in opposition to Russett, claims it is not democracy that matters, though both see the im-

¹² Britain and France Face off in the Fashoda Incident. *History Channel*. Retrieved from <https://www.historychannel.com.au/this-day-in-history/britain-and-france-face-off-in-the-fashoda-incident/>. Accessed 30 March 2019.

portance of getting the right concept in the claim. As Rosato (2003) argues: ‘Evaluating whether the democratic peace finding is caused by democracy or by some other factor such as American preponderance has implications far beyond the academy. If peace and security are indeed a consequence of shared democracy, then international democratization should lie at the heart of American grand strategy. But if, as I have suggested, democracy does not cause peace, then American policymakers are expending valuable resources on a policy that, while morally praiseworthy, does not make America more secure.’

Co-extensionality does not always cause problems for explanation or prediction of course. What matters is whether it breaks down or not in the targeted application. Does it? That raises a whole new set of questions. For instance, what is ‘American preponderance’? How is it measured? Where can it be relied on? To answer these takes a whole new set of research activities, many of which will look outward to new domains that have not yet played much of a role in DPT.

In Lecture 3 I shall argue that the credibility of DPT models and predictions comes from the virtuous tangle of interlacing practices that make up democratic peace theory—interlacing both internally and with those of other domains. (Thanks to Alison Wylie for the adjective ‘virtuous’.)¹³ These are all familiar activities, many well studied by philosophy of science, a few less so. What is important is to see them in one scheme, as *all together* what constitutes democratic peace *theory*—or better: *theorising*.

We philosophers of science have tended to have far too thin a notion of what makes for scientific theory, which results in far too narrow accounts of what makes the results of our theorising credible. Theory is not a set of propositions, nor a set of models, nor anything of that ilk. It is, as many from the Society for the Philosophy of Science in Practice advocate, a rich set of interlocking practices. The credibility of our scientific activities depends on the richness of this set and the interlacing of the activities that make it up—each done with a look over the shoulder to the others. In the remainder of these lectures I am going to talk about three kinds of activities that go into middle-range theorising that my own work has engaged with: Lecture 3 is on the use of Elster-style mechanisms, Lecture 2 is on theories of change and causal-chain modelling and in the remainder of today’s lecture I am going to look at what it takes to design a good measure.

DESIGNING A MEASURE

I begin with what we philosophers call ‘precisification’.¹⁴ As we use the terms in everyday language, both ‘democracy’ and ‘war’ are too loose to play a proper role in science. Many social science concepts seem to refer to specific qualitative or quantitative features that things might have, or to sets or functions of these. The number of women refugees in a province, the average age of school leavers, the NAIRU (the non-accelerating inflation rate of unemployment), or even just the ‘simple’ unemployment rate are examples. Other concepts sort things into categories based on a loose set of criteria where the members of the same category do not share any specific set of features but may rather have what Wittgenstein re-

¹³ Alison Wylie, Conversation, November 2, 2018, Seattle, Washington.

¹⁴ On the analogy of found art, Sophia Efstathiou calls these ‘found’ concepts that need to get ‘founded’ properly into a science if they are to play a proper role there. See Efstathiou (2012).

ferred to as ‘family resemblance’. Concepts with relatively clear evaluative implications, such as human welfare, human rights, the quality of PhD programmes or the number of people in poverty, generally fall in this category, along with many ‘everyday’ concepts at play in the social sciences. ‘Democracy’ and ‘war’ seem to be among these.

I follow Otto Neurath in calling concepts like this *Ballungen* (‘congestions’),¹⁵ as in the German “Ballungsgebiet” for a congested urban area with ill-defined edges: there is a lot packed into it; there is often no central core without which an item doesn’t merit the label; different clusters of features from the congestion (*Ballung*) can matter for different uses; whether a feature counts as in or outside the concept, and how far, is context and use dependent. Concepts like this don’t generally make for good science. They are too vague, too loose. To use them in a proper social science study they need to be made more precise, which is just what must happen to ‘democracy’ and ‘war’. But any concept familiar to general society, such as ‘disability’ or ‘poverty’ or ‘functional literacy’, is likely to be a *Ballung* concept with a multifaceted meaning. So in carrying these concepts into social science, we must be prepared for both multiplication of concepts and loss of meaning. To precisify concepts like these, we will likely have to sacrifice or alter large chunks of the meaning—what CG Hempel in commenting on Rudolf Carnap’s notion of explication has called ‘alienation’ of the concept: ‘An explication sentence does not simply exhibit the commonly accepted meaning of the concept under study but rather proposes a specified new and precise meaning for it’ (Hempel, 1952, p. 663) ... ‘At present, in fact, the connection between the technical and the pre-scientific meaning of theoretical terms has become quite tenuous in many instances, but the gain achieved by this “alienation” has been an enormous increase in the scope, simplicity, and experiential confirmation of scientific theories’ (Hempel, 1952, p. 701).

That brings us to the question of just what goes into *designing and defending measures*. This is a question I have worked on with Norman Bradburn, who is a pioneer in social indicators and quality of life research, and Jonathan Fuller, who is both a medical doctor and a philosopher of medicine (Cartwright *et al.*, 2017). Together we have developed a tripartite *theory of measurement*.

Measurement, we argue, is not just assigning categories or numbers; it is assigning values in a systematic and grounded way. It requires

1. A characterisation of the concept or quantity, which includes identifying its boundaries and fixing which features belong to it and which do not (*characterisation*).
2. A metrical system that appropriately represents the quantity or concept (*representation*).
3. Rules for applying the metrical system to tokens to produce the measurement results (*procedures*).

The reasons we undertake a measurement project —what we want to use the measurement results for— may affect one or more of these steps. Although I list them separately, what happens in each stage should influence each other stage. We may, for example, come to re-characterise a category on the basis of results derived relative to a candidate metrical

¹⁵ For discussion and references, see Cartwright *et al.* (1996).

representation of it. Or we may pick a metric system because the procedural rules for applying it are well-defined, or users know these methods better, or they are easier to implement. In the end the three components must fit coherently together. That's what 'representation theorems' are for: they are to show that the metrical system is appropriate to the concept as characterised and that the procedures can give accurate information about the concept characterized and in the form the representation demands.

Consider now precisification in the context of this theory of measurement. That the original Ballung concept has proven useful in many ways, the same may not be true for the more precise concept relative both to the way the world is and to what our purposes for it are. It may for instance lump together items that behave differently in ways we are concerned with or it may separate items that behave the same. Or it may lend itself to distorted understanding. For instance, yes-no poverty measures, as opposed to 'depth of poverty measures' have been criticised for incentivising governments to bring the richest of the poor just above the poverty line, ignoring the plight of those much worse off. This makes their poverty figures look better but it does not give an accurate picture of the amount of poverty people are experiencing and the amount of concomitant suffering.

Returning to the case at hand, let's look at some samples of how democratic peace theorising measures democracy.

Characterisation

How we define a democracy often depends on what functional emphasis is placed on a political system. Democratic theorists who focus on institutional or procedural structure are apt to define democracy in terms of various rights and freedoms or structural relations about how the government works, as in for example Robert Dahl (1971, p. 128):

- i. The right to vote.
- ii. The right to be elected.
- iii. The right of political leaders to compete for support and votes.
- iv. Free and fair elections.
- v. Freedom of association.
- vi. Freedom of expression.
- vii. Alternative sources of information.
- viii. Institutions that depend on votes and other expressions of preference.

A government is, or is not, a democracy depending on whether it meets these criteria.

Alternatively, the focus may be on outcomes so that the defining aspect of democracy is success at ensuring certain freedoms and standards to its people, rather than the system itself:

'Democracy' is one of the most used and contested concepts in social science. This theoretical diversity is amply reflected in the empirical realm. While definitions and operationalizations of democracy vary quite a bit, we can make a broad distinction between 'minimal' and 'maximal' definitions of democracy. The former focus on the importance of 'means', that is, procedures such as fair elections, respect for human rights, and universal suffrage. In contrast, maximal definitions include not only democratic procedures but also 'ends', or outputs (such as economic equality and social services). (Baviskar and Malone, 2004, p. 4)

Here is a selection of characterisations on offer. The point to note is how many and how varied — and these are only a *very* small sample.

- For modern states, democracy ... is usually identified with a voting franchise for a substantial fraction of citizens, a government brought to power in contested elections, and an executive either popularly elected or responsible to an elected legislature, often with requirements for civil liberties such as free speech. (Russett, 1994, p. 14)
- I define a liberal democracy as a state that instantiates liberal ideas, one where liberalism is the dominant ideology and citizens have leverage over war decisions. (Owen, 1994, p. 89)
- If a political regime is categorized as democratic only if the identities of the leaders of its executive branch and the members of its national legislature are determined in elections in which at least two independent political parties participate, in which at least half the adult population are eligible to vote, and if the fairness of elections has been established by at least one peaceful transfer of power between opposing political parties, then an examination of controversial cases reveals that the proposition that democratic states never fight wars against each other is defensible. (Ray, 1993, p. 251)
- ... government with the *consent* of the governed. This formula is indeterminate with respect to institutional forms, or the procedures by which consent is to be expressed. (Whelan, 1983, p. 14)
- Democracy is a competitive political system in which competing leaders and organizations define the alternatives of public policy in such a way that the public can participate in the decision-making process. (Schattschneider, 1960, p. 141)
- Government by the people, where liberty, equality and fraternity are secured to the greatest possible degree and in which human capacities are developed to the utmost, by means including free and full discussion of common problems and interests. (Pennock, 1979, p. 7)
- A study of Greek city states classifies ‘as a democracy any city that had an established assembly of the *demos* and in which a party purporting to represent the *demos* had some power’. (Russett, 1994, p. 46)

Representation

- The simplest kind of representation of ‘democracy’ is as a dichotomous variable: ‘yes’ if a country is a democracy, no if it is not. We would expect this when a list of criteria is set which must all be met, as in Dahl above.
- Alternatively an n-point spectrum might be indicated according to how many of the criteria are met, or how well.
- There are a variety of other more complicated ordinal ranking scales, like those composed by Polity IV and Freedom House below.
- Cardinal index numbers are also used, as in Vanhanen’s Index of Democracy which is on a scale of 0 to 100. (See below for how this is operationalised.)
- Fuzzy set representations have also been proposed.
- In the Greek City State study, to reflect both the nature of the city states themselves and of the data consulted, a 4-valued variable was used: Clear Democracies; All Other Democracies; Unknown; Not Democracies. Clear Democracies are those

judged to have met both criteria in the last characterisation listed above. The 'All Other' category consists of 'those that did not but for which some evidence of democracy exists.' (Russett, 1994, p. 47). This category is further coded into i) 'a democratic faction was in power', ii) 'an assembly was called for some constitutional decision (such as to declare war or assert autonomy) but there is no evidence of a regular assembly', iii) 'they are called democratic without evidence by Thucydides or modern scholars', iv) they 'were called a democracy but without convincing evidence or were colonies of democratic power, with some evidence of democratic practice'. (Russett, 1994, p. 47).

Procedures

Again, there are very very many of these, as we can see in Figure 3. I shall describe a small sample. We should note that the descriptions here are still fairly abstract, giving no idea about what is actually to be done on the ground to make the measurement, for example, to get answers to the questions in the Freedom House procedures. So there will be layers of work below this, with great numbers of choices to be made at each. Because of this layering, we should not try to draw any firm and clear boundary between what counts as *representation* and what counts as *procedures* in all cases.

— Freedom House¹⁶

- Measurements are made along two axes and split into a number of subcategories:
 - Political Rights
 - Electoral Process (3 questions),
 - Political Pluralism and Participation (4),
 - Functioning of Government (3)
 - Civil Liberties
 - Freedom of Expression and Belief (4 questions),
 - Associational and Organizational Rights (3),
 - Rule of Law (4),
 - Personal Autonomy and Individual Rights (3)
- Each question receives a score from 0 to 4, 0 being the least and 4 the most democratic.
- These aggregate scores are then converted to a number out of 7 for each axis, with 1 being the most free and 7 being not free. The average of the two determines the country's 'Freedom Rating', with (1.0-2.5) being Free, (3.0-5.0) being Partially Free, and (5.5-7.0) being Not Free.

¹⁶ Freedom House is a US-based 501 US-government funded non-governmental organization (NGO). Independent Auditors' Report. *Freedom House Inc.* Retrieved from https://freedomhouse.org/sites/default/files/FINAL_Basic_Financial_Statements_2016.pdf. Accessed 30 March 2019. Its methodology is found here: <https://freedomhouse.org/report/methodology-freedom-world-2018>. Accessed 30 March 2019.

- Scores are assigned by a group of analysts (publicly listed on the Freedom House website) in tandem with academic advisors.
- Polity IV¹⁷
 - Polity IV aims to study an aggregate of data from 1800-2017 of all independent states with more than 500,000 members.
 - It uses a 21-point scale ranging from -10 (hereditary monarchy) to 10 (established democracy), sometimes broken up into three sub-categories: Autocracies (-10 to -6), Anocracies (-5 to +5) and Democracies (+6 to +10).
 - The data only contain information on the central government and groups acting within the scope of the society. Break-away groups and other separatists not governed by the central authority are not considered.
 - ‘There is no “necessary condition” for characterizing a political system as democratic, rather democracy is treated as a variable. For example, the scale discriminates among Western parliamentary and presidential systems based on the extent of constraints on the chief executive. Charles de Gaulle as president of the French Fifth Republic operated within slight to moderate political limitations. Thus the early years of the Fifth Republic have lower Democracy scores than the United States or the Federal Republic of Germany, where constraints on the executive approach parity.’¹⁸
 - Scores are determined through two variables: ‘Institutionalised Democracy’ (DEMOC) and ‘Institutionalist Autocracy’ (AUTOC). A series of weighted scores are given to a set of questions for each category including ‘Competitiveness of Executive Recruitment’, ‘Constraint on Chief Executive’, and ‘Regulation of Participation’ amongst others. Scores for AUTOC are subtracted from DEMOC to give the final score.
 - Scores are given by a group of dedicated coders. Efforts are made for consistent coding practices, including having multiple coders work on the same data-sets, and sustained work to clarify the language used to formulate the criteria (what, for example, constitutes a ‘faction’ or a ‘transition’).
- Vanhanen’s (2017) Index of Democracy
 - Relies more heavily on data rather than the scoring of questions by analysts.
 - Uses two axes: ‘competition’ and ‘participation’.
 - Competition is defined as the percentage of votes for small or independent parties, or seats held by small or independent parties in government. It is calculated by subtracting the largest party’s percentage from 100.
 - Participation is defined by the percentage of adult citizens who voted.
 - The minimum threshold for a democracy is considered to be 30% competition, and 10% participation.
 - Both axes are combined to give the country’s Index of Democracy.

¹⁷ Center for Systematic Peace. Retrieved from <http://www.systemicpeace.org/polityproject.html>. Accessed 30 March 2019.

¹⁸ Center for Systematic Peace. Retrieved from <http://www.systemicpeace.org/inscr/p4manualv2017.pdf>. Accessed 30 March 2019.

And what procedures do you think were employed by Russett to categorise the Greek City states as he indicated?

ADDITIONAL DEMOCRACY AND GOVERNANCE MEASURES

BTI - Bertelsmann Transformations Index:

<http://www.bertelsmann-transformation-index.de>

Index developed by the Bertelsmann Foundation on the political and economic development status of 120 countries in transition (2003-2010).

Democracy Ranking:

<http://www.democracyranking.org>

Democracy index based on political (among others Freedom House, CPI) and socioeconomic factors. Available for 100 countries between 2008 and 2010.

Democratic Audit:

<http://www.democraticaudit.com>

Qualitative appraisal of the democratic quality and the human rights in several countries. Democratic Audit is an independent research group at the University of Liverpool.

Economist Intelligence Unit Democracy Index:

<http://www.eiu.com>

Democracy index developed on the basis of expert evaluations of 60 factors from five areas (election process, civil rights, government capability, participation and political culture) in 167 countries in the years 2006, 2008 and 2010).

Freedom House: Freedom in the World Reports:

<http://www.freedomhouse.org>

Ratings on the guarantee of political rights and civil rights in all countries worldwide. Data available from 1972 onward and yearly updated. Well-established rating developed by the American NGO "Freedom House".

NID: Neuer Index der Demokratie:

<http://www.politikwissenschaft.uni-wuerzburg.de/~nid>

New democracy measure developed by Hans-Joachim Lauth (University of Würzburg) based on the combination of constituent parts of the Freedom House, Polity and WGI indices. Data is available for 60 countries since 1996 on a two year basis.

Polity:

<http://www.systemicpeace.org/polity/polity4.htm>

Classification of political systems on a scale between the two extremes autocracy and democracy. Yearly updated data available for all countries worldwide from 1800 onwards. One of the most used democracy measures, it is compiled at the Colorado State University.

Polyarchy Dataset:

<http://www.nd.edu/~mcoppedg/crd/datacrd.htm>

Democracy index based on Robert Dahl's (1971) concept of polyarchy. Developed by Michael Coppedge (University of Notre Dame) and Wolfgang Reinicke (Global Public Policy Institute) the index is available for all countries worldwide for the time span 1985-2000.

SGI - Bertelsmann Sustainable Governance Index:

<http://www.sgi-network.org>

Index on the democracy level, the economic capability and the welfare state in 31 OECD countries. Developed by the Bertelsmann foundation, the index is available for the years 2009 and 2011.

UDS - Unified Democracy Scores:

<http://www.unified-democracy-scores.org>

Democracy measure developed by James Melton (IMT Lucca) and his team, it combines measures from 12 existent democracy measures (among others Freedom House, Polity, Polyarchy, Vanhanen). Available for all the countries worldwide between 1946 and 2008.

Vanhanen's Index of Democracy:

<http://www.prio.no/CSCW/Datasets/Governance/Vanhanens-index-of-democracy>

Democracy index developed on the basis of the polyarchy dimensions competition and participation proposed by Robert Dahl (1971). The index relies on election statistics worldwide (1810-200) and was developed by Tatu Vanhanen (University of Tampere).

V•Dem:

<https://v-dem.net>

Ratings on 11 different democracy components developed on the basis of evaluations by national experts. Available for all countries worldwide from 1900 onwards, the index was developed by the University of Gothenburg.

WGI - Worldwide Governance Indicators (Weltbank):

<http://info.worldbank.org/governance/wgi/index.asp>

Indices for six governance dimensions based on the combination of data from a broad array of data sources. Available for 213 countries in the time span 1996-2009.

Figure 3

Abdel-Fattah Mady (2014). Democracy Barometer. *Dr. Abdel-Fattah Mady's Webpage*. Retrieved from <http://abdelfattahmady.net/index.php/research/research-interests-/453-democracy-barometer.html>

IN CONCLUSION

I have introduced the theory of measurement here for two reasons.

1. I hope that it has given you a feel for what kind of thing goes under the label ‘middle-range theory’ and for the great variety of scientific activities that will be involved in any middle-range endeavour. Designing measures is a case of middle-range theorising par excellence. First off, we want concepts that will be both descriptively accurate and scientifically interesting, for instance by being linked in systematic ways with others or by being useful for prediction or manipulation. Then, how do we precisify them in a way that preserves both their scientific interest and some modicum of their ability to describe what we wanted to describe in the first place?

When it comes to the final product, it is essential that the three components of the measure mesh properly, so designing a measure is an iterative process. It is also essential that we be able to show that they do so. In the best of cases we do that with a *representation theorem* that proves that the representation is appropriate to the concept as characterised and by providing a series of empirical postulates to show that the procedures can give correct information about that concept. This generally involves importing knowledge and assumptions from well beyond the domain in which the concept is originally embedded.

What I hope you can see from this brief excursus into the vast work on measuring even just one concept in DPT is how much work —thick, detailed work— it takes, how complex it is and how intertwined the efforts must be, both internally within the measure itself and externally with work from other areas. This tangle of effort will play a big role in my final lecture.

2. I wanted to talk about middle-range theory in these lectures because I believe that middle-range theorising of all sorts is dramatically under-theorised in philosophy. Despite excellent recent work in history and philosophy of science and the earlier volumes by Duncan Luce and Patrick Suppes and associates, I think that this is still the case with respect even to this very central activity of designing and defending measures. So I offer here the framework developed by Bradburn, Fuller and me as a sample of the kind of work that philosophy can do on what goes on at the middle range.

Lecture 2. Causal-chain models and theories of change

THEORIES OF CHANGE

This lecture provides a philosophical account of the ingredients required for good causal-chain models and theories of change. Roughly, a causal-chain model lays out a series of significant steps, one after the other, by which a cause produces an effect. These models are often called ‘mechanisms’. To distinguish them from other senses of the term, I shall call them ‘causal-chain mechanisms’. The models I have in view are middle-range. They are theories of the causal pathways from particular inputs to particular outputs afforded systematically by structural mechanisms. Figure 1 is a good example. It is a casual-loop diagram explaining how a programme of monitoring, checklists and targets undertaken to improve

child welfare had negative consequences. It is taken from the *Munro Review of Child Protection* mentioned in Lecture 1.

Causal Loop Diagram exploring systemic impact of efforts to improve social work through increased prescription of practice

Developed in collaboration with Dr David Lane, London School of Economics and Political Science.

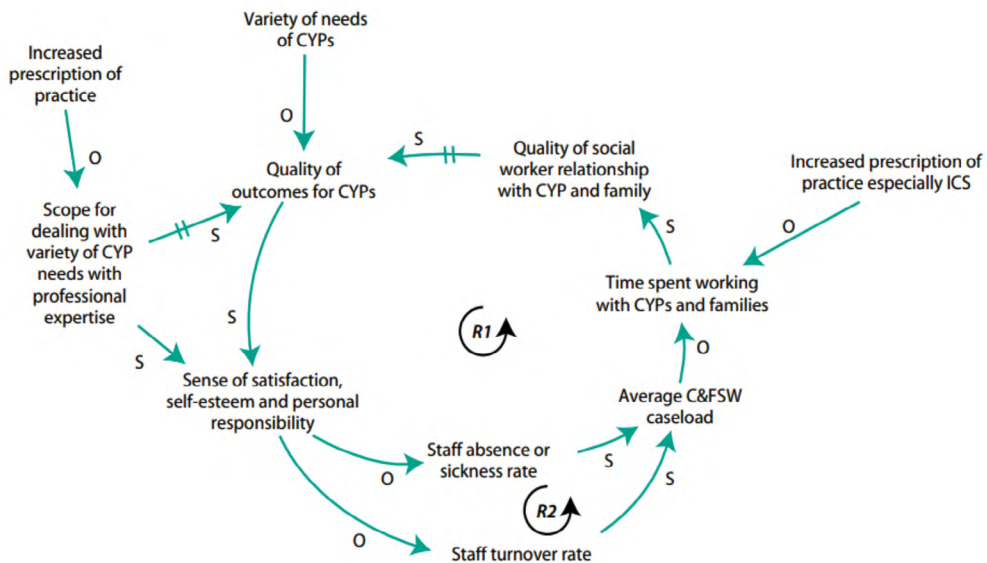


Figure 1

UK Department for Education (2011). *Munro Review of Child Protection*. Appendix. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/175391/Munro-Review.pdf

Theories of change, which are now widely demanded in policy contexts, are a subset of these middle-range causal-pathway models. Just as in Figure 1, in a theory of change the input is a policy or programme and the output is an outcome that the programme is supposed to achieve. Theories of change are sometime called ‘theories of action’ or ‘logic models’. Here is one typical description: ‘When properly constructed, logic models... convey the underlying theory of action—how the program’s resources, activities, and outputs lead to desired outcomes.’¹⁹ Here is another, relevant to my running example of the democratic

¹⁹ Wendy Kekahio, Brian Lawton, Louis Cicchinelli, and Paul Brandon. *Logic models: A tool for effective program planning, collaboration, and monitoring*. US Department of Education: IES, p. 3. Retrieved from https://www.researchgate.net/publication/261177247_Logic_models_A_tool_for_effective_program_planning_collaboration_and_monitoring

peace: 'A causal logic is a statement about how an independent variable exerts a causal effect on a dependent variable. It elaborates a specific chain of causal mechanisms that connects these variables and takes the following form: A (the independent variable) causes B (the dependent variable) because A causes x, which causes y, which causes B ... In the case at hand, democratic peace theorists maintain that democracy has various effects, such as support for peaceful norms of conflict resolution, which, in turn, increase the prospect for peace' (Rosato, 2003, p. 585).

You can see a catalogue of theories of change for the democratic peace in Rosato (2003 p. 586). One scholar's specific model is reproduced in Figure 2. For an example from a different domain, Figure 3 is the theory of change for the educational programme *Read 180* that I mentioned in Lecture 1.

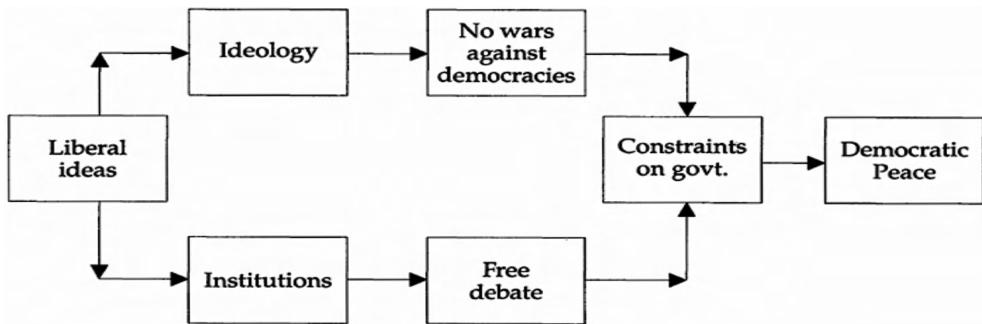


Figure 2

Owen (1994, p. 102)

This kind of step-by-step theory —with variables connected by arrows, 'A causes x, which causes y, which causes B'— is supposed to be helpful in deciding what programme to adopt, in figuring out how to implement it and in evaluating after the fact whether it did the job. This is true to some extent. You may look at the theory and realise that some of the requisite intermediaries will never be possible in your setting. And if you do decide to try the programme, knowing what is supposed to happen when allows you to monitor that the programme is on track. You can also use the information for *post hoc* evaluation. Suppose the targeted outcome occurs. Observing that each of the requisite intermediate steps obtained, and did so just when it was supposed to, provides strong evidence that the programme was responsible.

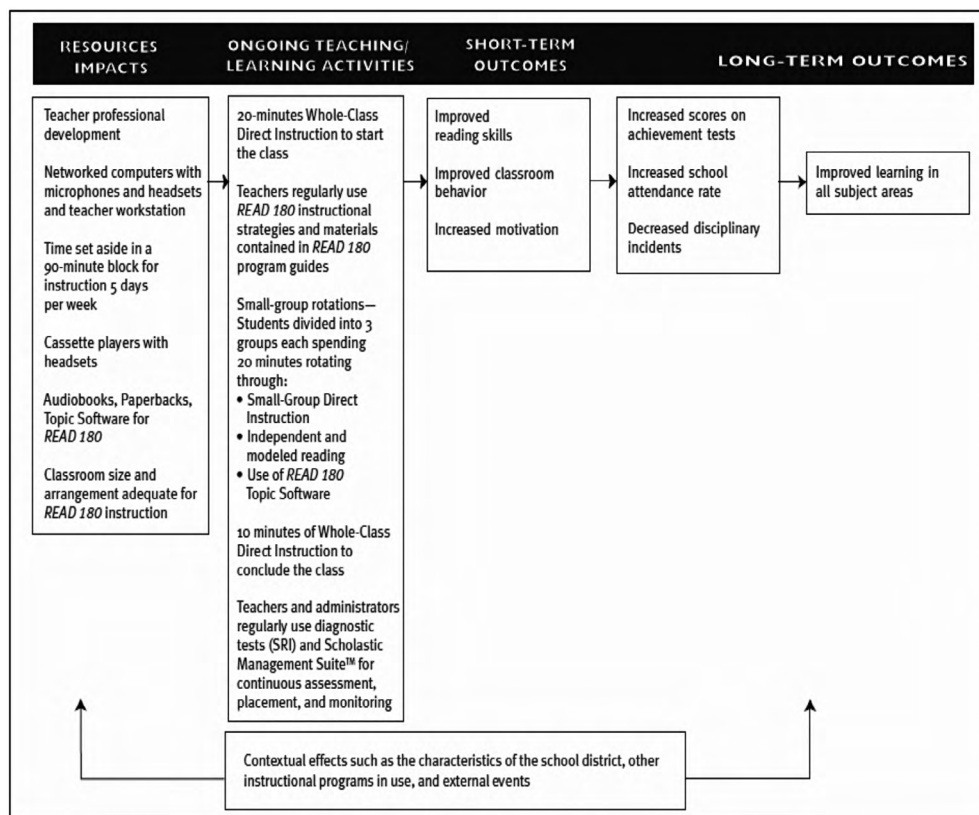


Figure 3

U.S. Department of Defense Education Activity. *READ 180 Research Protocol and Tools*. Introduction, p. 3.

Retrieved from https://www.scholastic.com/dodea/read180_administrators.htm

THREE PROBLEMS

There are three major hurdles to the effective use of theories like these. All three are raised among practitioners themselves and all three are rooted in solid philosophical work on causality. The first is highlighted in JL Mackie's INUS account. The second, which I shall argue is underwritten by a descendent of Wesley Salmon's (1984) process/interaction theory of causality, is a complaint we can find in literatures on process tracing and in what is called 'realist evaluation'. The third is the central objection of realist evaluation to step-by step theories of change. It depends on the same philosophic argument made in Lecture 1, that what will cause what depends on the underlying structural mechanism (which realist evaluation calls 'context', reserving 'mechanism' for something else).

Problem 1. No matter how full the programme design for policy interventions like Read 180 or deworming is, the policy intervention is seldom enough on its own to produce

the targeted effect. Policy programmes need what Jeremy Hardie and I call ‘support factors’ (Cartwright and Hardie, 2012) if they are to work. A theory of change that leaves out what support factors must obtain, or be put in place, to operate with the policy programme to produce the effect is seriously incomplete.

Problem 2. Policy and programme evaluation is a big industry: did this policy do what it was supposed to when implemented? will it work here? where will it work? Realist evaluation is one school of how to go about it. Realist evaluators are realist about the causal processes and activities by which policies produce their outcomes. They side with ‘mechanist’ social theorists in complaining that theories of change like those in Figures 1, 2, 3 and 4 ignore activities (the ‘mechanisms’ by which effects are produced) and picture only variables at the nodes, with thin arrows in between.

For instance, from Peter Hedström, who helped found the field of analytic sociology, and economic sociologist Richard Swedberg in their introduction to *Social Mechanisms*: ‘... the increasing use of [survey analysis and statistical techniques] has ... fostered the development of a variable centered type of theorizing that only pays scant attention to explanatory mechanisms’; and from leading expert on realist evaluation Ray, Pawson (2000, p. 301-303): ‘...in most survey and evaluation research, theory is ‘flattened’ so that it is expressible only in X à Y propositions ... Theory is indeed flattened so that middle-range questions about contexts, mechanisms are squeezed from the agenda’.

Theory is flattened in two ways. One of the things that is wrong with these ‘variable-and-arrow’ theories is that they have nothing to say about the underlying structure that affords these causal processes, which is the core of my third problem. The other is that they do not show what the cause at each step *does* to produce the effect at the next step. They represent the effects achieved by a cause but not the activities that the cause engages in in order to produce those effects. This makes for troubles in at least two ways.

- a) The first relates to support factors. If we do not know how the cause is to produce its effect, we will have a hard time figuring out what the requisite support factors are. Ray Pawson and Nick Tilley (1997) give an extended example of the reduction in car thefts by installing CCTV cameras in car parks. Do the cameras scare the thieves off from fear of being recognised? Or do they allow the police to arrest the thieves by dashing to the car park as soon as they see suspicious activity? Or in one or more of a dozen other ways they suggest. If it is the first, then one of the support factors is that the cameras be clearly visible to prospective thieves. If it is the second, that they are well hidden.
- b) Knowing what the cause does is almost essential for identifying those contexts in which it can do it, i.e. to identifying the structural mechanism that affords that causal pathway. For instance, as we shall see in more detail below, many explanations for the democratic peace suppose democracies do not go to war with other democracies because liberal ideas (whether in the populace as a whole or among liberal elites) on both sides take liberal ‘norms’ in other states as indicators of the legitimacy or lack of legitimacy of force against them. Does this mean that ‘democracy’ as characterized in either of the two ways below is not a good enough marker to pick out contexts in which serious tensions will be resolved short of war?
 - For modern states, democracy ... is usually identified with a voting franchise for a substantial fraction of citizens, a government brought to power in contested

elections, and an executive either popularly elected or responsible to an elected legislature, often with requirements for civil liberties such as free speech. (Russell, 1994, p. 1)

- Democracy is a competitive political system in which competing leaders and organizations define the alternatives of public policy in such a way that the public can participate in the decision-making process. (Schattschneider, 1960, p. 141)

Do we need instead ‘liberal’ democracy, as for instance with John M Owens: ‘I define a liberal democracy as a state that instantiates liberal ideas, one where liberalism is the dominant ideology and citizens have leverage over war decisions.’ (Owen, 1994, p. 89). With this definition, Owen argues, the marker does not catch in its net the US-British War of 1812, which has troubled scholars as a possible counterexample to the democratic peace, since ‘the War of 1812 was fought at a time when almost no Americans considered England a democracy.’ Owen (1994, p. 90).

Problem 3. This is the one from Lecture 1. Different causal pathways are possible in different contexts. That was the point about pressing the lever on the toaster versus on the toilet. So a causal-chain model will hold some places and not others. But there’s nothing in the boxes and arrows to indicate that.

PROBLEMS AND THEORIES OF CAUSALITY THAT BACK THEM UP

These are serious complaints that do not just follow from local arguments and examples in practice communities. Each of the three is backed by a serious philosophic theory of causality. Together the three corresponding theories of causality have real sting because they are mutually consistent and together they provide a rich account of how the causality represented in theories of change is supposed to work. Just what do I mean by ‘backing’ here? Here is an answer that Hagge Cadenas offered in commenting on this lecture:²⁰

First, one should note that when X backs up Y, X lends a sort of support to Y. If a strong muscular person is backing me up, then whoever is against me, risks facing the strong person. Similarly, if the theories are backing up the problems... whoever believes the problems are negligible pays the price of rejecting the philosophical theory. For example, a causal chain theorist who does not feel the need to include support factors risks paying the intellectual price of rejecting Mackie’s very intuitive INUS account [...].

... [Second], the theories also play a prescriptive role. This role is captured by [Figure 4]. How do they give such prescriptions? I think it is that the three philosophical theories are general theories of how causation occurs and hence how it should be represented: As Cartwright says: “... [the theories] provide a rich account of how the causality represented in theories of change is supposed to work”. Mackie says it occurs with support factors. Salmon says causes occur via processes and interactions. The realists claim that causes occur because of structural mechanisms and contextual features. The reason these theories of causation relate to theories of change (a sub-class of causal-chain models) is that changes occur because of causation... [T]heo-

²⁰ Hagge Cadenas (‘Cartwright on Good Causal Chain Models’. Dept. of Philosophy, UCSD, March 2019).

ries of change, then, are subject to theories of causation because the latter are philosophical accounts of the former...

[So] the three philosophical theories back up the three problems in two distinct ways. First, the philosophical theories place an intellectual price tag for whoever wishes to deny that the problems are genuine problems. Thus, one type of backing the theories do is motivational. Secondly, the philosophical theories, by being an account of causation/change, provide a general framework for how theories of change should be formulated. Thus, they back up the problems by playing a prescriptive role: one should formulate theories of change in such a way that avoid the three problems.

If you accept		Theory
Mackie's INUS account	then you should	Include support factors
Process-interaction theory	then you should	Include causal activities
Structural Mechanisms	then you should	Include context

Figure 4

By Haggeo Cadenas

I'll describe each of these three theories, developing them in a way that I hope makes clear their use in understanding theories of change and the phenomena these theories are supposed to represent. Since I introduced structural mechanisms yesterday, I can keep that discussion very short, so I will start there. Next I'll turn to support factors, where I believe the philosophic theory will be familiar. Last I will turn to processes and activities, which will take the bulk of my effort.

THEORY 1: STRUCTURAL MECHANISMS/ CONTEXT

What causal pathways are possible depends on the underlying structure. As Michael Strevens (2012) points out, sometimes it is left implicit what this structure is though the intended referent is recognised in the community of practice. Often it is picked out by markers, as I have illustrated with the democratic peace. In cases where we are especially well informed, a description of the relevant features of the structure can be supplied. But when a causal-chain model is offered for use, the mechanisms that afford it must be indicated in some way or another.

THEORY 2: SUPPORT FACTORS

Let us label the effects represented at each stage in the theory of change by X_i , as in Figure 5.

$$\begin{array}{c} X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_n \\ X_n c \equiv A_{n1} Y_{n1} \vee \dots \vee A_{nm} Y_{nm} \end{array}$$

Figure 5

JL Mackie (1965) taught that generally causes are INUS conditions for their effects: Insufficient but Necessary parts of Unnecessary but Sufficient conditions. It is usual to represent this thus, as in Figure 5:

$$\text{Mackie formula: } X_n c \equiv A_{n1} X_{n1} \vee \dots \vee A_{nm} X_{nm}$$

Here X_n is the effect at stage n and the ‘ c ’ in front of the equivalence sign is to signify that the features referred to on the right-hand side are its causes and not mere correlates.²¹ The X s are distinguished from the A s in this version of the formula in that the X s are meant to represent causes of interest. The A s are then thought of as the ‘support factors’ that must also be in place if that cause is to contribute to the production of the effect. In general A is to be filled in by a conjunction of a large number of factors.

This familiar Mackie formula is for cases involving dichotomous (yes/no) variables. For multivalued variables the analogue looks like this:

$$\text{Generalised Mackie formula: } x_n = \sum_1^m a_{ni} x_{ni}$$

In epidemiology it is common to represent a set of factors represented in a single term of a Mackie equation —factors that together are sufficient for a contribution to the outcome— as slices in a ‘causal pie’, as in Figure 6.



Figure 6

²¹ So this is not a theory of what causality IS but rather about the form of causal relations. In Mackie’s account the causality is constituted by a kind of *fixity*.

Although there may be many possible causal pathways that lead to the same effect, represented in the separate terms in the Mackie formula, the theory of change for a programme only describes the one generated by the programme. So at each step the other ways in which the outcome can be produced are ignored. This omission seems appropriate given that the job is to provide a theory of how the programme is to achieve the outcome, not a theory of everything that can contribute to that outcome.

More worrying is the omission of the support factors needed at each stage to produce the next. These are essential if the outcome is to be produced, so it seems that the theory cannot be of much use for planning or monitoring if it leaves these out.

To remind ourselves about the omission of these support factors, I recommend a notation like Figure 7, which replaces the nodes on the path with wedges to signify that we are representing at each stage only one element of the causal pie that it takes to produce the next — an element that the stage before played a role in producing. A more satisfactory theory of change would fill these in, as Figure 8 suggests.



Figure 7

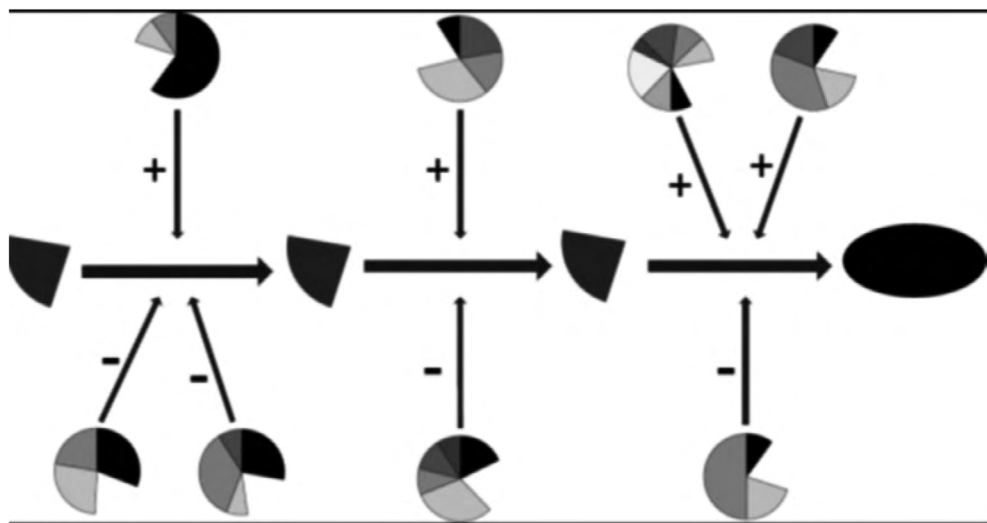


Figure 8

THEORY 3: ACTIVITIES AND THE PROCESS/INTERACTION THEORY

The claim that causality is not just one inert event after another but, rather, causes are active in producing their effects is familiar in philosophy. The philosophic theory of causality that I think fits best with the ‘realist’ and process-tracing practice recommendations I mentioned is one I build from Wesley Salmon’s ‘at-at’ process/interaction theory.

Salmon’s theory is an attempt to rescue the notion of causation from the damaging attack of Bertrand Russell. Russell was concerned with orderly recurring sequences of events like those described by the laws of physics: for instance, the elliptical orbiting of the planets around the sun, described by Kepler’s laws and derivable from Newton’s. However one wants to think about the relation between these laws and the sequence they describe — perhaps the sequences just happen to keep happening in the same orderly way and the laws just describe that, as in a Humean view, or our scientific laws represent laws of nature that somehow govern these happenings— Russell argued that there are special problems with the idea that one event in the sequence *causes* another later to happen. The problem is timing. The cause is gone before the effect occurs. How can it then do anything, let alone bring a whole new effect into existence?

Salmon’s account has two key ingredients: causal processes (which include enduring objects and systems, like particles and social institutions) and causal interactions. When an event at one time and place is supposed to cause an event somewhere else later, the at-at theory pictures the first event as a causal interaction of two (or more) *causal processes* that either initiates a new process or produces a change in one of them. The affected process is then located at every point between the cause event and the effect event. Some time before the effect, the causing process intersects or overlaps with the affected process and *causally interacts* with it: the first *does something* to the second.²² The effect process is thereby changed in such a way that it now, at the right time and place, possesses the characterising features of the effect event.

The theory is called the ‘at-at’ theory following Russell’s own solution to Zeno’s paradoxes of motion, adapted to the space-time relations between cause and effect. There is a gap between the events we call *cause* and *effect* but there is no puzzle about how the cause, which no longer exists in that gap, can produce the effect at the end of the gap. The cause event is an interaction (either instantaneous or stretching over a duration) of spatially and temporally extended continuous processes. The interaction changes one of them or initiates a new process. The initial cause is not a cause of some next event in the process. The changed process simply exists with the relevant changes (or as I will note, the naturally evolved versions of them) *at every point* up to and throughout the next interaction.²³

Although Salmon tried hard to use as few causal concepts in his theory as possible, it is clearly not a reductive theory, nor did he claim it to be. The account is riddled with causal discourse, especially as I have laid it out. Salmon himself stressed the importance of distinguishing *causal* processes, like a beam of light from a rotating beacon, from non-causal ones,

²² Perhaps the second also does something to the first.

²³ In commenting on this lecture, Ahmed Siddiqi asks ‘how does the process just exist at every point?’ (Personal correspondence. UCSD, 15 February 2019). The answer from the point of view of both Russell and Salmon is that it just does. That is what happens naturally in the empirical world. See also my discussion below about ‘natural evolution’.

like the moving spot on a stadium wall created by that beam. Only interactions with genuine causal processes can cause later effects: inserting a red filter someplace on the wall to turn the light spot there red cannot cause later effects in the moving spot on the wall.

To diminish the causal concepts employed in the analysis, Phil Dowe (2000) developed a special 'energy conservation/exchange' form of the process/interaction theory in which the interactions are exchanges of energy between energetic processes. Then the energy, being conserved, exists at every point in the process until the next exchange. The energy in the light beam is conserved from point to point: the beam can figure in causal interactions. The energy is then absorbed in the wall, appearing in some micro processes there. It does not appear in any successor spots on the wall: the 'moving spot' is not a process that can participate in causal interactions.

The hope here was that *causal* processes could be replaced by processes in which the energy is the same at each place and time and *causal* interactions could be replaced by reduction of energy in one process and increase of energy by the same amount in another at the same space-time point. I can't see how this can work. It seems possible for the energy in one process —process A— to transfer to a second —process B—, thereby diminishing by some amount, say ϕ , the energy in A, at the same time and place as the energy in process C is increased by that same amount ϕ from interaction with process D. The energy has not been *exchanged* between A and C so events in A cannot cause ones in C but the account seems to judge that it does so.

Whether this worry can be handled in the energy-conservation version of Salmon's process/interaction theory does not really matter for our purposes though. For it certainly doesn't help with causation in social settings. Still, I think the original Salmon ideas can help in the social sciences, so long as we do not try to use them to produce a reductive account of causation, which I at any rate have long argued is impossible. I want rather to use these ideas to develop an account not about what causality is but about what it is like, an account that makes sense of the 'realist' and process-tracing practice recommendations I mentioned. This involves developing Salmon's theory to make it more responsive to social science settings, which were not the kind of cases Salmon had chiefly in view. This in turn requires *inter alia* keeping firmly in view not only the processes involved but the settings in which they are to occur, which for our purposes are the settings to which the theory of change is supposed to apply —the structural mechanisms that afford them.

For these purposes I should note that although Salmon worked to distinguish causal from non-causal processes, he did not attempt to give identity conditions for processes, nor did he grapple with the issue of the natural development of a given process. The first of these —just what a process is— is an important question that is the subject of much serious work in metaphysics right now. I defer to this in order to discuss issues of more immediate concern for understanding theories of change. What matters for our purposes is that, whatever the metaphysics of processes is, a process should be able to sustain the features needed for the process/interaction theory.

The second issue requires more comment for our purposes. A process exhibits an array of features at different times. Which features obtain when and where are characteristic of the kind of process it is in the setting in which it occurs. For example, a golf ball hit off a tee in a St Andrews' golf course is a causal process, one that lends itself to the Dowe-style story. It changes its position over time, flying through the air and eventually striking the earth with a thud, crushing the grass beneath it.

This process/interaction story supports a typical philosophical example of a singular causal claim, geared to mesh with the ‘energy conservation’ version of the process/interaction theory: the launching of the ball in just that way at the tee caused the crushed grass in the St Andrews’ course (or, thinking of a causal prediction, ‘will cause the grass to be crushed there’). We may suppose, idealising, that apart from a minor effect of the drag of the air, the ball retains the energy imparted to it by the swing of the club.

What I have just described is the natural evolution of *that* process in *that* setting. The interaction of the ball with the golf club initiates the process, ‘the flight of the ball’. What is important about this flight-of-the-ball process vis-à-vis our causal claim is that it has features during the period of interaction with the grass that allow the ball to crush the grass. I’ve tried to design the case so that the feature of the flight-of-the-ball process that matters for the crushing is the energy, which as it happens has been the same throughout the process. But what makes our causal claim true on my more general version of the Salmon process/interaction theory is not that the features that matter have to do with energy nor that they are the same across the history of the process. Rather, it is that they are features that

- a) in combination with the features of the grass allow for the interaction of the ball and grass to crush the grass
- b) are features of the process that the interaction of the swing with the golf ball initiated and
- c) are features that the interaction with the swing introduced into that process.

There are several things to note.

1. There are no counterfactuals involved. Salmon did not like counterfactuals and was at pains to avoid them. On his theory, it is irrelevant whether, had the swing not initiated just that flight-of-the-ball process, some other process would have interacted with the grass to crush it. The concern is with what actually caused what, *in situ*; with what did the job — not what might have done the job if the current cause had not been on duty. This is one of the features that make the Salmon process/interaction theory suitable for understanding the causal accounts given by evaluators and causal-process tracers. Consider: it may be foolish to hire me to accomplish a result that will happen in some other way if you don’t hire me. But this does not excuse you from paying me if, on post hoc evaluation, it is clear that I did what I was supposed to, that what I did is what in point of fact produced the outcome.

2. The ball interacts with the air during the whole course of the process. This is definitely a fact about *this* process in *this* setting — the process that (we are supposing) actually happened (or will happen) at St Andrews’ course. The interactions with the air are irrelevant to the truth of our causal claim about what happened (or will happen) because they do not rob the process of the features it must have to interact with the grass. But other interactions along the way might matter. Generally, interactions that occur between the cause event and the effect event that change the evolving characteristics of the causing process can affect that process in different ways, which in turn affects the truth of the causal claim in different ways. If they remove features necessary for the final effect, the original causal claim will be false. In the opposite direction, they can enhance the process in two different ways. They can add features that by themselves are what actually bring about the effect without resort to those possessed before these interactions occur. Then the original interaction will not have causal responsibility for the effect. Or, recalling the facts about INUS causality, they can add

support factors —factors that cannot by themselves produce the effect but work with factors already in the process to do so. In that case they share causal responsibility.

3. *Causation* is always via causal *interaction*: no interaction, no causation. Between interactions, whatever is happening is not causal. So earlier events in a process do not cause later ones. In the simplest cases, like the idealised golf ball, where the features created at the initial interaction are the same ones that are essential for producing the effect during the later interaction, the features ‘get’ from the earlier interaction to the later by being *at* every point in between. That’s the point of Salmon’s at-at theory. For more complicated —more realistic— cases, features essential for producing the effect in the later interaction *evolve* from those produced at the earlier one. In those cases it is features that have evolved from ones produced at the start and that will evolve into ones essential for producing the effect later that are *at* the appropriate places *at* the appropriate times. For this to make sense, of course, we need to accept that there are natural processes of evolution for some features, described perhaps by natural laws, and that these are not processes in which earlier states cause the ones that evolve later. This is just what this feature does over time, much as we suppose that a body in motion tends to remain in motion —that is its natural evolution with no causal interactions necessary to get from earlier to later states.

Although causality is missing, there is no suggestion that it is arbitrary what features are where when. The evolving pattern of kinetic and potential energy in the golf ball in its specific setting in St Andrews is in accord with ‘natural law’. In the social science cases, the same is true for whatever is the appropriate analogue of natural law there. This should be true for any account of laws and principles in the natural and social sciences so long as the account does not suppose laws must have to do with causation. On the process/interaction account, *causation* —one thing *making* another happen— is something over and above things happening as scientific laws describe.

4. For Salmon, the cause is always an interaction or an event that marks an interaction, just as the effect is. But there does not seem to be any loss in weakening that requirement. On the at-at story as I have told it, there are features at every point in the process from initial interaction that evolve ‘naturally’ into those that are essential for the production of the effect in a subsequent interaction. I propose to count the appearance of features at the appropriate time and place that will evolve into ones essential to the ultimate production of the effect as causes as well.

5. Thin causal ascriptions should give way to thick ones.²⁴ Processes don’t just *interact*; they *do* things to each other, specific things. The CCTV camera does not just cause prospective thieves to go away; it *scares them off* or *motivates them* to try elsewhere. The golf ball thudding to the ground does not just cause the grass to lie flat; it *crushes* the grass.

Is there something to be said about what makes thick causal interactions *causal*? I think not, nothing very informative, and certainly nothing involving less ‘causally loaded’ language. The one process *changes* the other, and it does so in such a way that *produces* the features that designate the effect.

²⁴ For a nice discussion of the ‘causatives’ that express these thick concepts in natural language and their surrounding philosophic advantages, see Julian Reiss, ‘Cause, causatives and theories of causation’. Retrieved from http://www.jreiss.org/Presentations/HK_Cause%2C%20Causatives%2C%20Causation.pdf

Nor do I understand why we should want more. Hume enjoined us to look for something more because he couldn't find these causal processes nor their thick interactions in his highly filtered world view. If we have them in ours, what need is there for some further characterising features that fairly unambiguously sort interactions into two bins, causal versus non-causal? I may want to know that the CCTV camera did its job. I hired it to prevent car crime and car crime diminished. Did the cameras earn their pay or was this just a lucky outcome for them? If they convince me that they scared prospective thieves away, that's all I need to know. I don't need further assurance that 'scaring away' in this case is an instance of 'causing'.

I say this despite the fact that I have spent decades studying *causal* inference and methodologies for establishing *causal* claims. But what we are doing in establishing a causal claim, I take it, is establishing that there *is* a thick relation (or a set of relations) holding between the putative cause and effect that has certain features of interest in the case at hand—for instance, raising the probability of the effect once other 'causes' have been held fixed, which is what the probabilistic theory of causality finds interesting.

6. When the golf ball thuds to the ground, it is no accident that the result is some sorry looking grass. A heavy object thudding onto a fragile one is just the kind of thing that results in the fragile object being crushed. Although this may not be among the laws of our fundamental physics, this a familiar fact of nature, it is the way the world works. This matters when we are in the business of laying out a theory of change. We need reassurance that we can rely on the activities that are supposed to produce the targeted effects: they do so in accord with the way nature operates, albeit not as described in our most fundamental and general laws. As Jon Elster argues 'Specifying a causal chain does not mean giving up on general laws altogether, only going from general laws at a high level of abstraction to laws at a lower level of abstraction.' (Elster, 2015, p. 25). Much of the third lecture will be devoted to these lower or middle-range laws. For now I just want to insist that a good process/interaction theory should employ them. These are essential for showing that the activities we cite in our causal models can do what we say they do.

7. James Woodward has pointed out that 'the philosophy of science and structural equation literature has focused almost exclusively on type-level causal claims... By contrast, philosophers working in the tradition inaugurated by David Lewis have focused exclusively on token causation'. (Woodward, 2005, p.74). I take the process/interaction theory as I have expanded it here to be appropriate to both.

SO: I propose that a process/interaction model be adopted as a general framework for theories of change, on which ideally *a theory of change will involve four essential ingredients*:

1. The 'natural' (non-causal) evolution of factors from the place and time of the cause to that of the effect
2. The nature of the interaction by which the effect is produced — the 'activity'
3. The principles —often middle-range— which the interaction instances
4. An account of, or at least a marker for, the kind of underlying structures that afford the sequence of changes theorised

This accords well with what Derek Beach and R B Pederson (2019, p. 30) describe in their text on process-tracing:

...a causal mechanism can ... be portrayed as

$$X \rightarrow [(n_1 \rightarrow) * (n_2 \rightarrow)] Y$$

This should be read as X transmits causal forces through the mechanism composed of part 1 (entity 1 and an activity) and part 2 (entity 2 and an activity) that together contribute to producing outcome Y. This is a “context free” mechanism [NC: i.e. free of mention of the underlying structure that affords the surface-level causal mechanism pictured], and a proper study would also detail the contextual conditions that enable the [NC: surface-level] mechanism to be activated.

DEPLOYING THE THREE PHILOSOPHIC ACCOUNTS

With these three philosophical theories as tools, let us now consider how we can handle the three objections raised to theories of change like those represented in Figures 1-4. The first caution to note is that whether an objection is apt depends on the purpose for which the theory is intended. I have been looking at the different purposes to which such figures are commonly put: *post hoc* evaluation, *ex ante* prediction and middle-range theorising. Since our topic here is middle-range theory, I will focus on that. In this case the aim is to present an account that holds widely and reliably across some domain that is picked out either implicitly or explicitly, an account that is reasonable to take as default-correct failing reasons to the contrary (though, as always, with due attention to the chance of failure). How do the problems fare in view of the three philosophic theories about causality I have adumbrated?

Context. Theory 1 claims that whether the causal pathways pictured in the middle-range theory of change will obtain depends on the underlying structural mechanism. Without specifying what this is, the theory of change is dramatically under specified. We need to know *where* it will hold widely reliably since it certainly won't hold many places.

This information can be added as, for example, suggested at the bottom of Figure 3, but with two cautions.

- The description of the context actually has to be added, whereas in Figure 3 all that is done is to take note of the fact that context matters. In fortuitous cases, a complicated account of the structure that affords the pathways represented will not be necessary —the job can be done with markers. This, though, invites another stage of middle-range theorising, as I stressed in Lecture 1, to make credible that the markers are indeed reliable.
- Beware the arrows from the bottom level upwards. The horizontal arrows in these kinds of representations are standardly taken to mean that the feature at the head of the arrow causes the feature at the tip.²⁵ It is a matter of some discussion what these arrows could sensibly mean. Carl Craver (2007), for example, maintains that the iterations of the structural mechanism *constitute* the surface causal behaviour. John Pemberton, Sarah Wieten and I (Cartwright *et al.*, 2018), by contrast, argue that the features of the parts of the structural mechanism are governed by general principles, often familiar ones, and the surface behaviour just is what it is for the parts to obey those principles in the particular arrangement found in that mechanism. Whatever other account we might endorse we surely should not suppose that the operation of the structural mechanism *causes* the causal relations pictured in the fig-

²⁵ Of course it is the features *represented* at the head and tail that are supposed to be causally related. For simplicity of presentation I will not be careful about use/mention distinctions like this except in cases where it might be confusing to conflate the two.

ures. So we need some different understanding of them or other. I am going to go with the Cartwright/Pemberton/Wieten account because it makes place for principles, which play an essential role in ensuring that results of the activity in the effect-producing interaction are not happenstance.

Support factors

When the aim of the figure is to present a middle-range theoretical claim, I think my own objection here will often be misplaced even though it is based in the Mackie theory of INUS causality that I endorse. That's because we expect the figure to be relatively uncluttered and the level of specificity to be uniform across it. But generally, at the level at which the figured cause is described, there are often a great many configurations of support factors that can work with it to produce the effect. We philosophers may be prone to miss this since our usual example is that of striking a match to produce a flame, with the support factor of oxygen in mind. There's not much you can substitute for oxygen if you want to start a fire. But there often is in social settings. Homework will not produce increased learning without a number of support factors, but some are intersubstitutable. For instance, if pupils do not have a quiet supportive place to work at home, an afterschool homework club may do as well.

Of course one can always deal with this by moving up in abstraction. But often there's no informative description to give beyond, 'add the right thing at the right time,' which should go without saying. Alternatively, one can fill in some particular set, or sets, of support factors at the right level of specificity. But this can be misleading — suggesting that there is good theoretical reason for supposing these particular features are necessary, as with the oxygen — and can also correlatively curtail imagination about what might do as well.

This argues for regularising a convention with respect to these kinds of theoretical representations — a convention that I suspect is actually already often in place, though practitioners may not articulate it. I propose then that a good rule to adopt for this kind of middle-range theorising is that

- it should be understood that the cause depicted requires support factors if the effect is to be achieved
- in settings where it is to be applied, either one or more sets of sufficient support factors should obtain or be reasonably easy to put in place
- specific support factors are represented if, but only if, they are necessary widely across cases (like oxygen for the flame)
- in any particular setting, for either *post hoc* evaluation or *ex ante* prediction, more detailed information with full sets of support factors needs to be provided.

What does the cause do?

What shall we do about the objection that the answer to this question is not generally at all apparent in theories of change? A combination of my reconstructed Salmon process/interaction account with the account of the role of general principles in the operation of structural mechanisms provides an answer. Begin with the process/interactions theory, especially with the requirement that if C at t is to cause E at t' later in a particular kind of setting, first, C must comprise features whose natural evolution in that kind of setting are

essential in the interaction at t' to the production of E there, and second, those evolved features must *do* something in that interaction to produce E . This is not just a philosopher's demand. It is one that practitioners care about, and with good reason, because it can have real practical implications.

The same is true with the democratic peace theory. Since these are philosophy lectures, it is natural to start with Kant. In very brief: Kant believed that rational citizens will be against wars since they are both personally dangerous and economically costly. Thus, a government dictated by the will of rational citizens will be reticent to engage in conflict:

Now, in point of fact, the Republican Constitution, in addition to the purity of its origin as arising from the original source of the conception of Right, includes also the prospect of realizing the desired object: Perpetual Peace among the nations. And the reason of this may be stated as follows. According to the Republican Constitution, the consent of the citizens as members of the State is required to determine at any time the question, 'Whether there shall be war or not?' Hence, nothing is more natural than that they should be very loath to enter upon so undesirable an undertaking; for in decreeing it, they would necessarily be resolving to bring upon themselves all the horrors of War. And, in their case, this implies such consequences as these: to have to fight in their own persons; to supply the costs of the war out of their own property; to have sorrowfully to repair the devastation which it leaves behind; and, as a crowning evil, to have to take upon themselves at the end a burden of debt which will go on embittering peace itself, and which it will be impossible ever to pay off on account of the constant threatening of further impending wars. On the other hand, in a Constitution where the Subject is not a voting member of the State, and which is, therefore, not Republican, the resolution to go to war is a matter of the smallest concern in the world. (Kant, 1795)

Casting the basic ideas of this explanation into the process/interaction framework, 'democracies' can be re-described as states that do the will of rational agents. Such states *choose* peaceful resolution in times of conflict. In terms of our contemporary rational choice theory, given that the state will do the will of the rational agents that make it up, we can see this course of action as falling in one fell swoop under the two general principles 'Rational agents choose the action that maximises their expected utility' and 'The disutility of the danger and cost of war is (almost) always greater than the disutility from peaceful alternatives'.

'Choose' may seem a very thin causative that does not go far towards underwriting the demand for the description of the activity by which the effect is produced, perhaps hardly thicker than 'cause' itself when the casual agents are reasoning conscious people and their governments. But in this context I think not. In contemporary social science using the term 'rational' invites a placement in rational choice theory, where 'choice' is a highly constrained concept. We 'choose' among a set of pre-determined options about which we have 'preferences', which must satisfy a number of conditions that are demanding enough to introduce the concept of *utility*. 'Choose' also places the account on the normative side of the norm/structure divide I am about to describe.

A Kantian-style account is only one of a great many explanations of the democratic peace. DPT is thus a good testing ground for how well the process/interaction scheme I advocate serves as a model for middle-range theories of change. Here is another explanation, this one on the institutional as opposed to the normative side, again highly caricatured in order to make clear the basic form. The story is that when the two opposing states are democracies, voters in each see the other as powerful because democracies tend to be power-

ful and to have powerful allies. Voters don't like long wars and they definitely don't like losing, so they don't like wars with powerful opponents. In a democracy, leaders are responsive to the likes and dislikes of voters for fear of losing office. So the leaders find some other resolution than war.

We can graph this in a step-by-step variables/arrows diagram like Figure 5. But the process/interaction requires us to add activities and principles. So: given that voters' *dislike* losing and they *believe* that it will be hard to win against a powerful opponent, the voters *choose* to be opposed to war with this opponent (they are not, for example, coerced into this attitude). This outcome is not happenstance. It is an instance of the *principle*: 'People who choose freely and rationally choose their preferred option among those allowed', since, given that the political structure is a democracy, opposition to the war is allowed. The leaders *fear* losing office; the voters' opposition *motivates* them to hunt a peaceful resolution. That the attitude of the voters has this effect is not happenstance either. Their attitude can make a difference because of the *general principle*: 'Leaders tend to do what it takes to stay in power', and in a democracy doing what the voters like is what it takes.

Note that Figure 5 does not graph support factors. That's because, as I noted, for the explanatory chain represented there, the support factors may differ from case to case, with nothing very informative in common across them. That, as I argued, is one of the reasons that the activities matter. Knowing what the activities are can help us to identify what the support factors might be in particular cases. For instance, if voters are to favour a peaceful resolution to the conflict with another democracy because they prefer not to fight long, costly wars, they will need to recognise that this opponent is a democracy and that democracies tend to be powerful and have powerful allies. That can be aided by a lively free press and a culture in which the voters are politically engaged. Or by a strong and active Quaker movement that is quick to draw public attention to the dangers. Or by a variety of other local arrangements. But just what those local arrangements are will be different for different explanations involving different activities.

IN CONCLUSION

Theories of change are helpful for ex ante prediction, *post hoc* evaluation and middle-range theorising. But not in their usual form. To do the jobs they are employed for, they should be process/interaction theories of the kind I have developed here. Many thanks to Wesley Salmon for the start on this.

Lecture 3. Mechanisms and middle-range laws

EXPLANATIONS AND MECHANISMS

Although the maxim 'Democracies do not go to war with democracies' is widely accepted, as we have seen, there is no widespread agreement about its explanation. And explanations matter, as scholars engaged in DPT note:

- From Russett (1994): '[N]o merely empirical relationship can be compelling without a powerful theoretical explanation' (p. 30). Russett later explains that there are in the literature two chief 'causal mechanisms of peace.' (Russett, 1994, p. 84)

Table I. Glossary of Definitions of “Mechanism”

Bennett and George (1997, p. 1): “the processes and intervening variables through which causal or explanatory variables produce causal effects.”
Bhaskar (1979, p. 15): “the construction of an explanation for . . . some identified phenomenon will involve the building of a model, utilizing such cognitive materials and operating under the control of something like a logic of analogy and metaphor, of a mechanism, which <i>if it were to exist and act in the postulated way would account for the phenomenon in question.</i> ”
Boudon (1998, p. 172): “A SM [social mechanism] is, in other words, the well-articulated set of causes responsible for a given social phenomenon. With the exception of typical simple ones, SMs tend to be idiosyncratic and singular.”
Cowen (1998, p. 125): “I interpret social mechanisms (defined in greater detail below) as rational-choice accounts of how a specified combination of preferences and constraints can give rise to more complex social outcomes.”
Elster (1989, p. 3): “nuts and bolts, cogs and wheels – that can be used to explain quite complex social phenomena.”
Elster (1998, 45, his emphasis): “Roughly speaking, mechanisms are <i>frequently occurring and easily recognizable causal patterns that are triggered under generally unknown conditions or with indeterminate consequences.</i> ”
Gambetta (1998, p. 102): “I take ‘mechanisms’ to be hypothetical causal models that make sense of <i>individual</i> behavior. They have the form, ‘Given certain conditions <i>K</i> , an agent will do <i>x</i> because of <i>M</i> with probability <i>p</i> .’ <i>M</i> refers either to forms of reasoning governing decision making (of which rational choice models are a subset) or to subintentional processes that affect action both directly (as impulsiveness) or by shaping preferences or beliefs.”
Goldthorpe (1998, p. 18): “some process existing in time and space, even if not perhaps directly observable, that actually generates the causal effect of <i>X</i> on <i>Y</i> and, in doing so, produces the statistical relationship that is empirically in evidence.”
Harré (1970, pp. 101, 102, 104): “the structures, states and inner constitutions from which the phenomena of nature flow. . . the permanent or enduring conditions under which a certain kind of phenomenon will occur.” “The inner constitutions, structures, powers, encompassing systems, and so on, of which natural generative mechanisms are constituted, and of which the connection between cause and effect usually consists.”
Hedström and Swedberg (1998, p. 11): “Mechanism-based explanations usually invoke some form of ‘causal agent’ that is assumed to have generated the relationship between the entities observed.”
Hedström and Swedberg (1998, p. 13): “Mechanisms . . . are analytical constructs that provide hypothetical links between observable events.”
Hernes (1998, p. 74): “A mechanism is a set of interacting parts – an assembly of elements producing an effect not inherent in any one of them. A mechanism is not so much about ‘nuts and bolts’ as about ‘cogs and wheels’ – the wheelwork or agency by which an effect is produced. But a mechanism or inner workings is an abstract, dynamic logic by which social scientists render understandable the reality they depict.”
Keat and Urry (1982, p. 30): “causal explanations require the discovery both of regular relations between phenomena, and of some kind of mechanism that links them. . . . In describing these mechanisms and structures we will often, in effect, be characterizing the ‘nature,’ ‘essence,’ or ‘inner constitution’ of various types of entity.”
King <i>et al.</i> (1994, p. 85): “Some scholars argue the central idea of causality is that of a set of ‘causal mechanisms’ posited to exist between cause and effect. This view makes intuitive sense: any coherent account of causality needs to specify how the effects are exerted.”
Kiser and Hechter (1991, p. 5): “A complete explanation also must specify a mechanism that describes the process by which one variable influences the other, in other words, how it is that <i>X</i> produces <i>Y</i> .”

(Continued)

Figure 1

Citation: James Mahoney 2001.

‘Beyond Correlational Analysis: Recent Innovations in Theory and Method’,
Sociological Forum 16, 3, pp. 575-593.

Stable URL: <https://www.jstor.org/stable/684726>.

Accessed: 31-03-2019 09:34 UTC. Appears on page 579

Table I. (Continued)

Koslowski (1996, p. 6): “A causal mechanism is the process by which a cause brings about an effect. A mechanism is a theory or an explanation, and what it explains is how one event causes another.”
Little (1991, p. 15): “A causal mechanism, then, is a series of events governed by lawlike regularities that lead from the explanans to the explanandum.”
Schelling (1998, pp. 32–33): “I propose . . . that a social mechanism is a plausible hypothesis, or set of plausible hypotheses, that could be the explanation of some social phenomenon, the explanation being in terms of interactions between individuals and other individuals, or between individuals and some social aggregate.”
Somers (1998, p. 726, citing Coleman, 1986:1328): “meaningful connection between events as the basic tool of description and analysis.”
Sørensen (1998, p. 240): “My definition of ‘mechanism’ is simple: It is an account of how change in some variable is brought about – a conceptualization of what ‘goes into’ a process.”
Steinmetz (1998, pp. 177–178): “Within critical realism, a law is not a constant conjunction of events but the characteristic pattern of activity, or tendency, of a mechanism. More specifically, real structures possess causal powers which, when triggered or realized, act with natural necessity and universality as generative mechanisms.” [Accompanying footnote: “Generative mechanisms are ‘tendencies’ rather than ‘powers’ because they are not just potentialities but potentialities that may be exercised without being manifested.”]
Stinchcombe (1993, pp. 24–25): “As I use the word, <i>mechanism</i> means (1) a piece of scientific reasoning which is independently verifiable and independently gives rise to theoretical reasoning, which (2) gives knowledge about a component process (generally one with units of analysis at a ‘lower level’) of another theory (ordinarily a theory with units at a different ‘higher’ level), thereby (3) increasing the suppleness, precision, complexity, elegance, or believability of the higher-level theory with excessive ‘multiplication of entities’ in it, (4) without doing too much violence (in the necessary simplification at the lower level to the make the higher-level theory go) to what we know as the facts at the lower level.”
Stinchcombe (1998, p. 267): “I have defined mechanisms before as bits of ‘sometimes true theory’ or ‘model’ that represent a causal process, that have some actual or possible empirical support separate from the larger theory in which it is a mechanism, and that generate increased precision, power, or elegance in the large-scale theories.”
Tilly (in press): “mechanism-based explanations aim at modest ends: selective explanation of salient features by means of partial analogy. . . . Mechanisms are events that alter relations among some specified set of elements.”

Figure 1 (continuation)

Citation: James Mahoney 2001.

‘Beyond Correlational Analysis: Recent Innovations in Theory and Method’,
Sociological Forum 16, 3, pp. 575–593.

Stable URL: <https://www.jstor.org/stable/684726>.

Accessed: 31-03-2019 09:34 UTC. Appears on page 579

- From Owen (1994): “That we do not really know the causal mechanism behind the democratic peace means we cannot be certain the peace is genuine. It may be an epiphenomenon, a by-product of other causal variables such as those suggested by realist theories of inter-national politics.” (Owen, 1994, p. 88)

Note the expression ‘causal mechanism’ by both authors. Talk of ‘mechanisms’ in these kinds of explanatory contexts is routine. But what is meant by ‘mechanism’ here? That’s a good question, with lots of candidate answers. Look at Figure 1 from a review article in a sociology journal. It lists 24 different definitions, and, although some philosophers are there, this list does not include our structural mechanists, Bechtel, MDC or Glen-

nan,²⁶ all of whom had published on mechanisms by then, nor Mario Bunge (1997), who spoke at the Conference on Social Mechanisms at the Royal Swedish Academy of Sciences in 1996.

MECHANISMS MACH 3

I promised at the start of these lectures to sort out from this morass three distinct senses of ‘mechanism’ that I see at play in contemporary philosophy and in the human sciences. Two we have looked at already: mechanisms in the sense of underlying structures and mechanisms in the sense of causal-chain models, both of which, I have argued, are essential ingredients in a good mid-range process/interaction theory.

Consider again the theories of the democratic peace discussed in Lecture 2. Many of these are obviously mechanistic theories in the step-by-step sense. Mechanism in the structural sense is also implicated: ‘democracy’, I have suggested, works as a marker for a pair of countries whose structures make the causal pathway from tension to war unlikely; also central structural features of democracies are pointed to in the labelling of the intermediate steps in the causal pathways postulated by explanatory theories of the democratic peace. The third sense is there as well when activities are described and ‘mechanisms’ are called into play to stand behind them. I call these last mechanisms ‘middle-range law’ mechanisms because they figure as general precepts in description, explanation and prediction in the human sciences in much the way I see ‘laws’ functioning widely in science. This Lecture is dedicated to trying to make sense of them and to understanding their usefulness.

Consider familiar calls for an account of the mechanism *by which* the cause produces its effect. For instance, from Hedström and Swedberg’s ‘Introductory Essay’ to the collection, *Social Mechanisms*:

Assume that we have observed a systematic relationship between two entities, say I and O. In order to explain the relationship between them we search for a mechanism, M, which is such that on the occurrence of the cause or input, I, it generates the effect or outcome, O. ... a mechanism can be seen as a systematic set of statements that provide a plausible account of *how* I and O are linked to one another. (Hedström and Swedberg, 1998, p. 7)

And from Gary King, Robert Keohane and Sidney Verba’s classic text *Designing Social Inquiry*:

Some scholars argue the central idea of causality is that of a set of ‘causal mechanisms’ posited to exist between cause and effect. This view makes intuitive sense: any coherent account of causality needs to specify *how* the effects are exerted. (King *et al.*, 1994, p. 8)

For a good theory the answer to *how* takes a special form. As I urged in Lecture 2, just describing the activity is not enough. A good theory should also provide an account of *why that activity can produce the targeted outcome*. This presupposes that the outcome is not an

²⁶ Recall Lecture 1.

idiosyncratic result of the activity. Rather it is ‘recyclable’: there is some regularity and systematicity to it. As Jon Elster argues in his widely-cited account:

Roughly speaking, mechanisms are *frequently occurring* and easily recognizable causal *patterns*... (Elster, 2015, p. 26)

Still, there seem to be two big troubles with the claim that Elster-type mechanisms can function as precepts in social science.

First, as Elster says, mechanisms of this kind may be useful for explanation, but they are generally short on descriptive and predictive power, which are hallmarks of good science:

- They are too *malleable*. They are too unconstrained about which situations they cover and which not.

As Elster himself says, continuing from the previous quote:

...that are triggered under *generally unknown conditions*... (Elster, 2015, p. 26)

If the conditions that trigger a mechanism are unknown, whenever the expected outcome does not obtain when a mechanism obtains it seems all too easy to attribute that to a failure of triggering. There seems to be no way to police such excuses. So these kinds of precepts seem to be unfalsifiable and whatever conclusions we draw using them to be unworthy of trust.

Second, despite their malleability:

- They don’t stretch far enough. They don’t even cover all the situations they are supposed to explain.

In fact they may sometimes cover nothing at all. There is all too often no pattern to be found. As Elster continues, they also frequently have

...indeterminate consequences. (Elster, 2015, p. 26)

One reason for this is that a mechanism seldom operates on its own. We sometimes pretend they do when trying to describe their ‘natural’ outcomes. But that makes no sense. Mechanisms don’t operate in some kind of Platonic heaven ‘all by themselves’. They operate in real settings. And in real settings other real things have influence as well. So the characteristic behaviour we associate with a mechanism is often not what we should expect to happen when the mechanism operates.

I shall argue that neither of these problems bar these middle-range law mechanisms from the halls of scientific knowledge once we embrace —as we should— a *community-practice centred instrumentalist* account of scientific laws, or at least of this type of scientific law. We should not try to cash out Elster-type mechanisms as (hopefully true) claims from which we could, with additional (true) premises, derive explanations, predictions and descriptions. They are generally not claims at all. They are rather precepts, formulae, maxims that *we may learn how to use* to build claims about what happens in nature.²⁷ This, I

²⁷ In commenting on this Lecture, JC Gonzales (‘Mechanism & Middle-level Laws’, Dept. of Philosophy, UCSD, 2019) draws an extended comparison between my claim here and the views of John Haugeland (2017).

maintain, is true for a good many of our scientific laws across both the natural and the social sciences. I develop this view in detail in *Nature, the Artful Modeler* (Cartwright, 2019). I mention it here only to suggest that the social sciences may not be peculiar in this respect. But it is not necessary for middle-range laws in social science to work like laws in physics, biology and chemistry. What really matters is whether the basic ideas of community-practice centred instrumentalism can be developed to show how middle-range law mechanisms in social science can be both credible and useful for scientific explanation, description and prediction.

The centrality of community learning and community practices is underlined in the US National Academy of Science's [NAS] 'Definition of Science' (from their 2008 booklet, *Science, Evolution, and Creationism*):

Scientific knowledge and understanding accumulate from the interplay of observation and explanation. Scientists gather information by observing the natural world and conducting experiments. They then propose how the systems being studied behave in general, basing their explanations on the data provided through their experiments and other observations. They test their explanations by conducting additional observations and experiments under different conditions. Other scientists confirm the observations independently and carry out additional studies that may lead to more sophisticated explanations and predictions about future observations and experiments...

Because observations and explanations build on each other, science is a cumulative activity. Repeatable observations and experiments generate explanations that describe nature more accurately and comprehensively, and these explanations in turn suggest new observations and experiments that can be used to test and extend the explanation. In this way, the sophistication and scope of scientific explanations improve over time, as subsequent generations of scientists, often using technological innovations, work to correct, refine, and extend the work done by their predecessors.²⁸

Note the *interplay* of activities: correcting, refining, extending. This claim by the NAS is not news. Nor is it philosophically controversial. The process of mutual adjustment of the ingredients to make a stable structure is well-known. But we have little in the way of philosophic accounts of it, or of the ingredients that go into it or of how it supplies epistemic security. Philosophy has tended to focus on one kind of fit or other at one point of contact or other, most notoriously between a well-formulated hypothesis with well-defined concepts and either an experiment or a body of data. This leaves out way too much of the diverse tangle of practices that go into making successful science.

The work I report here is part of a new project with Jeremy Hardie and Eleonora Montuschi (Cartwright et. al., forthcoming) aiming to show how the right kind of tangle makes for trustworthy science. The central question is not whether these practices are important to and characteristic of science but rather 'Do they provide sufficient epistemology to make our science credible?' Hasok Chang offers a serious case that they do,²⁹ at least as I interpret him. His focus is on the coherence of the practices, which we agree is important. We think much else matters as well in the nature of the tangle itself. What makes for credibility, we

²⁸ <https://www.nap.edu/catalog/11876/science-evolution-and-creationism>. Accessed 31 March 2019, p. 10.

²⁹ Chang (2017) and watch also for his forthcoming book.

are so far suggesting, is a network of practices —what we shall call a ‘virtuous tangle’³⁰— that have a ‘sufficient’³¹ mix of several key features.

To be *virtuous* the network of practices should be

- a) dense —there should be a great many of them.
- b) diverse —they should be the ‘right’ mix where what is right depends on context; they are not, for example, all just studies (or even worse, studies of the same kind).
- c) tangled —they should be mutually scrutinising, as we have seen in Lecture 1 in the simple example of the three components that make up a measure.
- d) long-tailed —they should have lots of connections with practices far from their core.
- e) able to accommodate a large amount of otherwise established knowledge (both knowledge that and knowledge how).

This is clearly a big claim. Here I want to borrow these ideas to defend the scientific legitimacy of middle-range law mechanisms. Specifically I want to argue that middle-range law mechanisms use loose language and they can almost never be recast as true claims. Nevertheless they can be credible if they are embedded in a dense tangle of highly contested diverse middle-range theorising practices with long tails that accommodate knowledge and practices otherwise deemed acceptable. This is still a big claim, one that I will not hope to nail down here. But I do aim to get a start defending it by illustrating it with the case of the DPT. I hope at least to show that this is a promising programme for safeguarding the scientific status of Elster-type middle-range law mechanisms.

I shall turn first to the worry about the truth of middle-range law mechanisms. But that discussion will soon bleed into an examination of the worries about their loose language since it is the thick, mutually supporting tangle of contested practice that I believe can solve both.

Before the philosophy, I should make an historical note. As Hedström and Swedberg note, ‘It was Robert Merton who brought together the idea of mechanism with that of middle-range theorising ... Merton firmly rejected all attempts to develop general systems of sociological theory and advocated instead that sociological theory should deal with “social mechanisms.” The point is to locate a middle ground between social laws and description, Merton said, and “mechanisms” constitute such a middle ground.’ (Hedström and Swedberg, 1998, p. 5-6).

INSTRUMENTALISM: WHY NOT?

Often we talk as if the mechanism is something there in a setting that is responsible, in tandem with other factors, for what happens. Sometimes this may be fine. For instance, many of the mechanisms that Elster describes or that are deployed in social science explanations seem like familiar dispositions that we ascribe to individuals or institutions. But when we try to understand this in detail we are faced with a host of vexing questions that plague metaphysics —What is a disposition? Is it grounded in a counterfactual conditional? Can institutions really have them or only individuals? And so on. I think we can avoid going into these kinds of questions. To do so, I propose to move from the material mode to the

³⁰ Thanks to Alison Wylie for the name.

³¹ What counts as sufficient will be context-specific and is not an all-or-none matter.

formal, to characterise Elster-type mechanisms not in terms of things in the world but instead in terms of middle-range laws or precepts and the accompanying practices that use these to help describe and predict things in the world.

On a conventional realist view of laws this won't make a difference. The laws we use to predict phenomena are supposed to cover those phenomena: (descriptions of) the phenomena follow from those laws. And in principle these laws are supposed to be (at least approximately) true. So those laws can't be right where the associated patterns fail.

I urge that they can indeed be right. They just can't be true. This should be no surprise. Hardly ever do we find the kind of thing we call a law or a principle expressed in propositional form. This is just what we should expect given the view I want to defend that they are not claims at all. They are rather tools, tools we deploy, along with a whole kit of others, to build concrete models that can provide claims about real happenings. This is in accord with what Mauricio Suárez, Towfic Shomar and I argued in 'The Toolbox of Science'. (Cartwright *et al.*, 1995).³²

That important laws of science are frequently not true is a familiar theme from me. It is the meat of my first book, *How the Laws of Physics Lie*. After that I lost interest in scientific realism, becoming more concerned with questions about causal inference, modelling and how we use the laws of physics, independent of whether they are true, for instance in designing and building a laser. But the work on lasers, which I pursued with Suárez and Shomar in our LSE project on modelling, reinforced my early views that laws are not claims from which we infer models. We don't infer, we *build*. We use the laws in ways an extended community of scientists and engineers have learned and have trained each other to do. There is good, experienced practice, trial and error, maxims, rules of thumb. But there are few rules about how to use them. The laws, Suárez, Shomar and I argued, are tools in the toolbox of science. Good, reliable models are built by scientists and engineers experienced in the use of these tools, often employing modules they couldn't construct themselves and that are built by other practitioners experienced in the use of other tools.

That's an instrumentalist view. No question. And I think there is no trouble defending it as the best view around about laws and principles, across the sciences. But I was reluctant to embrace it entirely. For bad reason. In my youth I was introduced to the 'inference ticket' version of instrumentalism, and I couldn't shake this off. There are several reasons to avoid this metaphoric label. 'Inference ticket' sounds like a license or permission to go from A to B by inferring B from A. But:

1. That suggests you can read the ticket and recognise the destination. Yet these laws play only a small part in fixing what B is to be.
2. 'License'/'permission' is the language of rules. But there are few rules, though there are many honed practices.
3. The concepts the laws use —like *priming* or *implicit bias*— and the specific functional forms — Gm_1m_2/r^2 — matter to how they are deployed. It is hard to find place for this on the inference ticket view.

³² It also dovetails nicely with the 'Context-mechanism-outcome' frame of realist evaluation defended in *inter alia* by Pawson and Tilley (1997) and growing out of the 'critical realism' of Bhaskar (2008). But note that what realist evaluators call 'context' generally covers both what in these lectures I call structure and what I have labelled the 'support factors' and 'derailers' for the mechanism given that structure.

4. There is no A to start from. It is not some body of claims that makes our models credible. It is the tangle of finely honed, critically contested practices in the hands of skilled practitioners.

Recently I have returned to work on these issues, partly stirred by Hasok Chang's pragmatic realism and the work of Philip Kitcher and others that has brought pragmatism alive in our field again. But also —what matters here— by a drive to make sense of this widespread use of 'mechanisms' in social science explanation and prediction. What was missing from my youthful worries about instrumentalism was sufficient attention to the interlocking practices that go into deploying these instruments.

We know what to do with our well-tested principles to build good predictive and descriptive models —that is part of what it is to be well-*tested*. And, with respect to worry 3, what that is depends heavily on the notions and formulae the principles employ. I use the word 'know' intentionally. How these principles are used to build successful models is not accidental nor haphazard. It develops and evolves (in its own way, case-by-case). The way these notions and principles are used is generally complex, learning how to do it takes a lot of careful, detailed training, often imagination as well, and training in one set of application domains does not generally equip one to apply the same principles in others. Sometime things go wrong. But they mustn't do so too often or the principles and surrounding practices will need overhauling. Although the boundaries are vague, there will be clear ad hoc uses. When these are successful, it is generally accepted that more work needs to be done, developing the network of theory and practice to bring them into the corral or to explain why this ad hoc move works when it seems it shouldn't.

EXPLANATION AND PREDICTION

Thus far I have been playing fast and loose with the notion of explanation. I don't think the points I want to make require any specific account of scientific explanation, but there is one distinction that I think is important here. It is in line with Karl Popper's objections to the theories of Freud and Marx, which he maintained could explain almost anything. A plausible explanation of an individual's behaviour is offered, but had just the opposite behaviour occurred, the same theoretical principles could have been invoked to explain it, just filling in a different story. There is nothing in principle wrong with that from a scientific point of view. I can use the theory of gravity to explain both a pin's falling and its rising by filling in further facts about the presence or absence of a magnet. The trouble from Popper's point of view is that, with the theories he did not like, what further information could be added and just what role it could play was dramatically under-policed within Marx's and Freud's theories. Explanations like that are often called 'just-so stories'.

What about explanations with Elster-type middle-range law mechanisms? Many of these will be just-so stories. We can often imagine details that could be added to the narrative that could couple with the same principle to explain a contrary outcome. Note that here the issue is not that there may be opposing mechanisms that point to contradictory outcomes but rather that one and the same mechanism can (help) account for opposing outcomes.

I urge that we follow Popper here in opposing just-so stories in science. For an explanation to be *scientifically credible* it has to be scientifically justifiable as an explanation. We

have to have evolved practices, that together, and together with the facts we can learn about a case, are sufficient for us to say: 'Yes, that is a legitimate application of this mechanism'. The practices must be sufficiently developed to give a verdict and that verdict should not be highly contested in the case at hand.

Theories that do not have enough methods of constraint built in to provide scientifically credible explanations will also suffer when it comes to prediction. As Christopher Hitchcock and Elliott Sober put it: if a theory is 'sufficiently plastic that it can accommodate any data that may come along, it is in no position to make predictions about what data will come along.' (Hitchcock and Sober, 2004, p. 7)

Philosophy of science has long been concerned with issues about the symmetry of explanation and prediction. That is not being assumed here. To require that a *post hoc* scientific explanation employing specific mechanisms be justifiable as an explanation is not to require that that explanation could have allowed us to use the same mechanisms to predict the happening explained before the fact with enough non-prescient information, for at least a couple of reasons:

- Many mechanisms are probabilistic or need triggering, and very often the dispositions or arrangements supporting many different mechanisms that can affect the same outcome are in place, including ones that point in opposite directions, as Elster stresses. We can't tell beforehand which will operate. But there can be ample evidence afterward about that.
- As noted already, few outcomes are the result of the operation of a single mechanism. So even when we have a detailed understanding of a mechanism, and even when we are able to predict when it is likely to operate, we may be in a poor position to predict what will happen because we do not have the same deep understanding of the other significant mechanisms that will have an effect. This is well illustrated in economics where the theory surrounding specific mechanisms can be honed and honed until it seems the mechanism is well understood; nevertheless, it is useful for prediction only in special circumstances where it is the dominant mechanism at work. (For examples of economic mechanisms see Angus Deaton (2010), which also illustrates how mechanisms get corrected, honed and refined.)
- Even when we have a good understanding of the different significant mechanisms that will operate in a situation, we often lack nice 'rules of combination' sufficient to predict what happens when they operate together. But again there can be ample evidence afterward to trace how they operated together to produce what happened.
- A very great many income/outcome pairs we study are separated by a long sequence of steps in between, as we saw in discussing causal processes in Lecture 2. Each step must obtain if the final outcome is to be achieved. But each will depend on different mechanisms, with often (as above) more than one likely to operate at any given step, and each mechanism will have different support factors necessary for it to operate. This makes prediction difficult, whereas post hoc explanation may well be possible.

Let's look at this in DPT.

DEMOCRATIC PEACE THEORISING

The idea of the democratic peace has been around since the 1960's (Babst, 1964). It became big business in the 1990's and continues to be so today.³³ Work on the democratic peace involves thousands of studies and thousands of academic papers. In Lecture 1 I noted that democracy might be used as a marker for structures that make war an unlikely outcome of conflict. I think it is right to reconstruct its use as that of a marker and not, for example, an Elster-type mechanism, for a couple of reasons. First, DPT explanation schema like those represented in Figures 2 and 3 in Lecture 2 involve a number of successive stages, with different middle-range law mechanisms supposed to operate at different stages. Second, DPT advocates argue that the democratic peace holds almost exceptionlessly. This suggests that they are treating 'democracies don't go to war with democracies' as a claim —albeit a very loose one until the concepts are all well-defined and quantifiers inserted. This does not fit readily with the idea that democracy is a mechanism for peace, given that the 'natural' outcome of a mechanism is seldom what we can expect to happen when that mechanism is present, due to the action of other mechanisms, the possible failure of triggers and of support factors and the chance of derailers. This fits more naturally if we take 'democracy' as a marker —find a democracy and you will find a structure in which conflicts with other democracies get resolved short of war— in this case, as advocates would have it, a highly reliable marker.

Is it a reliable marker? Is it true that democracies don't go to war with democracies? In what sense of democracy? Of war? Why think the association is not just accidental? DPT has produced hundreds of explanations of individual cases that are supposed to fall under it or may seem deviant. Why trust these, and which ones? Though I don't look at any of this here, a vast amount of critical work has gone into investigating each of these issues.

I did however look briefly at another strand in the DPT tangle: samples of each of the three components of measures for the concept 'democracy'. Recall the need to get these three components to mesh. Here we are in familiar philosophic territory; namely, the demand for a theory of the instrument: how do we know that our instruments do what they are supposed to? Think then of the vast network of further work —sub-theorising— that must go into arguing that practical on-the-ground procedures that are offered can deliver the right information about the concept to be measured, especially since so many of our procedures look at indirect indicators postulated to have calculable connection back to the concept of interest.

Developing measures is just one among a myriad of interlocking activities that go into theorising the democratic peace. Here is a catalogue of different types that I have observed there:

- Development of new concepts (that social scientists sometimes call 'constructs')
- Arguments to establish construct validity for the concepts employed
- Development of measures for these concepts —many different ones
- Development of 'bridging' theories to defend that these are good measures
- Ensuring there is theoretical and empirical support for these bridging theories, which often involves information from different, sometimes distant, domains

³³ For example, Rosato (2003); Farnham (2002); Hegre *et al.* (2018); University of Gothenburg: Varieties of Democracy Institute: <https://www.v-dem.net/en/news-publications/working-papers/>; Weisiger and Gartzke (2016); Choi (2016) and Tomz and Weeks (2013).

- Development of well formulated hypotheses
- Testing of alternative hypotheses
- Statistical analyses, especially correlations and partial correlations of a vast variety of different sorts
- Case studies
- Considerations of which case studies are the ones to pursue
- Process tracing
- Historical and archival work (e.g. looking at newspapers, correspondence, official records)
- Ethnographic studies
- Experiments³⁴
- Postulation of new middle-range law mechanisms
- Building models and explanation schema
- Figuring out how to implement strategies suggested by these models
- Developing hundreds of individual case explanations
- Lots of connections with work outside the democratic peace theorising community

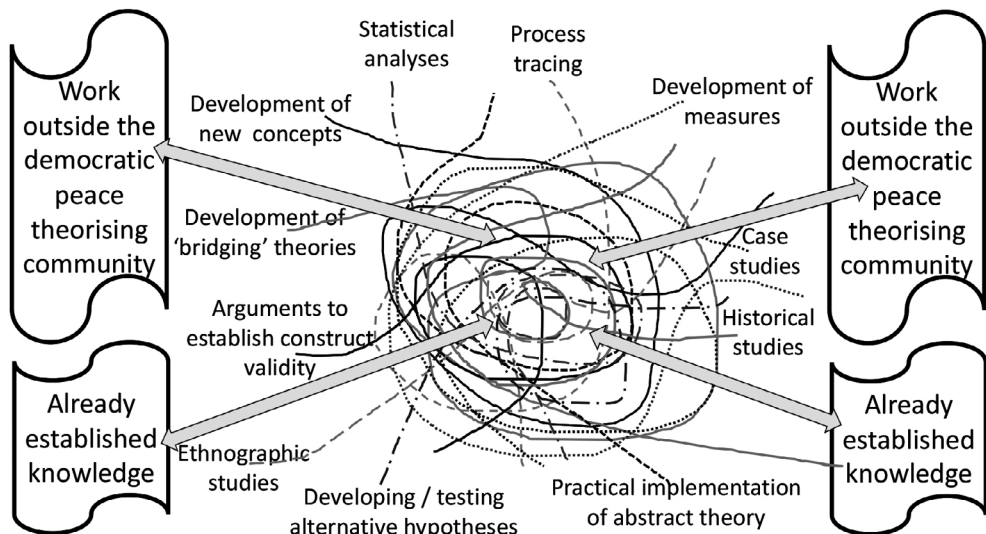


Figure 2

Constructed by John Pemberton

I put these in a list here. But for the credibility of the DPT practices and of explanations constructed with them, it matters that they are thickly interwoven and mutually supporting and constraining, as in the tangle in Figure 2. In these Lectures I have given only one example of this kind of mutual support and constraint, in Lecture 1 where I illustrated how

³⁴ Since experimentation can seem unlikely for this subject matter, I provide a few references: Tomz and Weeks (2013) and Geva and Hanson (1999).

the design of a good measure requires the right meshing of characterization, representation and procedures. If you look at the list above, I think it will be immediately plausible that the results of what each of these practices can constrain will have implications for many others. But showing how that works in detail in the case of DPT is research still in progress, so I will not venture further on it now.

What I will do here in defence of middle-range law mechanisms is now to tackle directly my worries about loose language.

WHY LOOSE LANGUAGE NEEDN'T MATTER

It seems I can readily be accused of inconsistency. I began Lecture 1 with a complaint about evidence-based policy: 'My second shock in engaging with EBP was with the exceedingly sloppy language employed throughout. According to conventional EBP wisdom, by careful review of the research evidence we can learn "What Works" ... But: what on earth could "It works" mean?' I have lectured about this both to philosophers and to EBP communities, driving home the point with the UK World War II slogan, 'Loose talk costs lives'.

Yet here I endorse middle-range law mechanisms not formulated as proper propositions and employing loosely characterised concepts. There is a difference though. There are maybe a hundred different definitions of democracy on offer in the DPT community, most with appropriately matched measures. Their use for different purposes and in different contexts is heavily critically debated, as is the correctness of the central DPT precept given various definitions of democracy, in tandem with various definitions of the concepts that go into those definitions (like defining 'competitiveness of executive recruitment'), in tandem with the same for 'war' and its defining concepts, in tandem with different ways of testing the extent to which any of these 'precisified' claims are true or could be relied on for different purposes in different contexts.

Nothing like this is the case for 'works' in 'It works'. Recall Read 180, where the What Works Clearing House reports 'Strong evidence of a positive effect [on comprehension] with no overriding contrary evidence.'³⁵ In cases like this the WWC will have provided a fairly precise characterisation of what both 'Read 180' and 'comprehension' consists in. It will also be clear about the criteria for 'strong evidence of a positive effect with no overriding contrary evidence' and you would be able to bore down and find out which studies are supposed to have provided that evidence.

That is all admirably precise, just what we should expect in credible science. But what about after that: 'It works'. In this case 'it' is Read 180, for which we have a precise meaning. It's the 'works' bit that is troubling, and it remains so even if we replace 'It works' by 'Read 180 has a positive effect on comprehension'. This is every bit as flabby and generic as a good many social-science middle-range law mechanisms, for instance, cognitive dissonance: 'People seek consistency among their cognitions and feel discomfort when it is missing'.

But that, in my view, need not be a problem. In both cases they could be tools to use in building concrete descriptions, explanations or predictions for real situations. The problem is that in the Read 180 case we have little know-how, no good, thick tangle of practices surrounding the development of predictions and explanations, whereas by now we do with the demo-

³⁵ https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_read180_112916.pdf

cratic peace, and, I suspect, with cognitive dissonance as well. Read 180 is not totally lacking thicker work though. It comes with a theory of change, which is why I chose it as an example in the first place, and its developers have made some efforts to connect it with various learning and psychological theories. Many other programmes that make it into approved lists in What Works sites do not have even this to recommend them as tools for planning and prediction.

The conclusion I want to draw from this discussion is that the fact that a middle-range law mechanism is loose does not mean that its use is uncontrolled. In the 1950s and 60s, philosophers concerned about Popper's charge of falsifiability distinguished the internal principles of a theory from the bridge principles. The bridge principles serve to police the application of the theory to the world. But we don't need principles to do this job. Principles fall in the category of knowledge-that. We can equally do it with knowledge-how, reflected in a thick tangle of critically evolved practices.

GETTING EPISTEMIC SECURITY

I propose that there are three elements that underwrite the credibility of explanatory models and predictions developed with the use of middle-range law mechanisms.

The first is reflected in the National Academy of Science's focus on the interplay of activities that I quoted at the start: '[T]he sophistication and scope of scientific explanations improve over time, as subsequent generations of scientists, often using technological innovations, work to correct, refine, and extend the work done by their predecessors.' What is involved in 'correcting, refining and extending'? I propose that these be understood in terms of developing a virtuous tangle. So I urge that:

1. The development of the model should be embedded in a network of mutually supporting practices satisfying my five criteria a)-e).

The second comes from the US National Research Council's testimony to the US Congress on findings from the NRC 2002 report, *Scientific Research in Education*, which is at pains to suggest that Congress should not be legislating scientific method:

The NRC report makes clear that the objectivity and progress of scientific understanding in any field...derives not from a given methodology or a given person. Rather, it comes from the community of researchers.... who embrace high standards of quality. (Taken from Eisenhart and Towne, 2003, p. 35)

The importance of the structure and habits of the research community is common fare in Science Studies and there is now a great deal of work in epistemology and in philosophy of science (from Helen Longino (1990) and others) on what about the community makes for the objectivity of its results. I refer you there for more discussion. I propose to follow their lead and add as a second major contribution to its credibility that

2. The work is carried out in a research community that works to 'high standards of quality' with an eye to each other's practices in order to 'correct, refine and extend' the community know-how.

The third is a tenet of everyday common-sense epistemology that I think works equally well in science: for reliable description, explanation or prediction, stick to the tried-and-

true. Consider a familiar contrast. I differ from my neighbour. He usually buys the newest, latest models of new goods; I, the boring old-fashioned ones. Why? I hate fussing. I want something that does what I have learned to expect of it the first time, every time. He is excited by novelty, wants the things he invests in to be fun and to do what he could not have done otherwise. For the excitement of this he is willing to put up with glitches, maybe lots of them. We have a difference in attitudes to adventure and to bottom-line functionality, but *we do not have an epistemic difference*. We agree that my purchase is more likely to do the good old tired-and-true than his and that his has a real chance of that new next step, which mine does not. It is the well-rehearsed tried-and-true practices, the next model as close as possible to ones already established as highly successful, that have the most chance of doing what we expect of them. It is the tangle of experienced practices that provides epistemic security. I trust that my device will work because of the web of repeatedly successful practices that has produced it. So:

3. The closer a predictive, explanatory or descriptive model is to ones that have been honed and tested and that have proved repeatedly successful in cases *like* this, where judgments of likeness are themselves backed up by a thick tangle of honed practices, the better. The further away, the more reason to hedge our bets.³⁶

Admittedly we often cannot stick to the tried-and-true. We face new problems with new aspects to them — we need new solutions. My claim then is not that it is always best to stick to the tried-and-true but rather that the further we venture away from it the more epistemically insecure our productions.

WHERE HAS ALL THE EVIDENCE GONE?

We philosophers are in the habit of thinking in terms of well-formulated hypotheses and specific pieces of evidence for these hypotheses. That seems to play little role in my story. That is not entirely true though. Return to the case of my more adventuresome neighbour and myself. He hazards that the new device will work as promised on the basis of the evidence the manufacturer adduces for it. It is not that I fail to look at their claims, nor that I think these claims are false,³⁷ nor that I fail to understand the force they are supposed to have. It is just that *evidence* is not enough to provide the high level of epistemic security that I, unlike my neighbour, demand. Nor am I interested in any specific well-formulated claim. It is all too easy for manufacturers to pick some specific precisifications of this claim that can be well evidenced for products that are not really fit for purpose. I want instead to be assured there is a large tangle of practice supporting the loose precept, ‘the new device will do all I expect of it’.

It may seem that here again I am being grossly inconsistent. I have spent a great deal of time in the last few years developing and defending a theory of evidence, the ‘Argument Theory’. I have offered it as a theory of evidence, ‘For policy and wheresoever rigour

³⁶ As always this is ‘*ceteris paribus*’ advice, and of course we can always be mistaken about how close the model is or about how similar in the right respects the case at hand is to ones where the model has worked.

³⁷ Though I often suspect that, that is not my point here.

is a must' (as the title of my monograph puts it (Cartwright, 2013)). And surely rigour is a must in making policy predictions, or in any other areas where people's welfare is at stake.

Yes. I agree. But the overall effect of the Argument Theory should be to make us cautious in our claims to have evidence for our predictions. On the Argument Theory, there is no fact of the matter about whether a fact *e* is evidence for hypothesis *H*. Evidence is not a 2-place relation but a 3-place relation. In the third place sit the reasons —the additional premises— that show why *e* is relevant to the truth of *H*. This is a version of the familiar need for auxiliary assumptions. Relative to one set of assumptions *e* can be evidence for *H*, and relative to another, that same *e* can be evidence for $\neg H$. For instance, Marilena DiBucchianico (2009) tells the story of two warring camps with opposed theories of high-temperature superconductivity. Each, because of their different background assumptions, took a newly discovered 'kink' in the experimental data to be evidence in favour of their own theory and against the opponent's.

So, where does the tangle of practice enter? Generally, the argument that establishes that *e* is relevant to *H* will involve a great many auxiliary assumptions and new concepts. *e*'s claim to relevance is no more credible than the union of these assumptions and the validity of all the concepts involved. In happy cases, at least some of these may already be 'established' or 'well understood' —we are entitled just to assume them. But that is not the norm. Whether we have evidence at all for our hypotheses is up for grabs then. To make the relevance claims credible, we need to render the assumptions credible and the concepts valid. And we need further credible assumptions and probably further new concepts to assure the relevance of the evidence adduced for the earlier ones.

The trouble I want to raise here is not that this could be a bottomless regress but rather how dicey it makes *e*'s claim to relevance. It doesn't take much to upset it. One response is to note that there may be many routes —many different sets of auxiliaries— which if true would render *e* relevant to *H*. We could suppose that at least one is likely to be true. Another strategy is to notice that we may have other results that look to be evidence for *H* as well. We could suppose that at least one of these is likely to be relevant. But what supports these suppositions? If evidence is required for credibility, we should have evidence for that, genuine evidence, not wishful thinking. And remember, all this is just about *e*'s relevance to *H* —does it speak for *H*'s truth?— before we begin to consider how loudly it speaks. My worry is that if we stick merely to the evidence game, we will have a hard time genuinely providing warrant for our claims.

Consider a claim involving Read 180, which requires expensive software and the time and money for staff training.³⁸ Suppose you are a California middle-school principal with a lot of children in your school who have recently immigrated from China and are reading

³⁸ The WWC Intervention Report says, 'As of January 2017, the initial start-up cost of a *READ 180*[®] *Universal* package for 60 students was approximately \$43,000. Houghton Mifflin Harcourt (HMH) provides 1 day of in-person professional development, a 2-hour webinar, and eLearning courses with the purchase of the program. A *READ 180*[®] *Universal* upgrade kit for 30 students costs \$11,000 and includes teacher materials, two HMH Teacher Central licenses, 30 *Real Books*, six boxes of Independent Reading Library books, access to the new online student application, and 30 HMH Student Central licenses. An upgrade kit with 60 student licenses costs \$15,000. There are also upgrade and full package options available for classes of 15 students. https://ies.ed.gov/ncee/wwc/Docs/Intervention-Reports/wwc_read180_112916.pdf. Accessed 2 April 2019.

at least two years below grade level. What warrants the prediction that Read 180 will bring their comprehension up to grade level? You will be told to look to the WWC for help. The WWC gives a ++ rating for Read 180 with respect to adolescent comprehension.³⁹ That rating is based on 6 studies, where in their own words, ‘In the six studies that reported findings, the estimated impact of the intervention on outcomes in the comprehension domain was positive and statistically significant for two studies, one of which meets WWC group design standards without reservations, positive and substantively important for one study, and indeterminate for three studies.’⁴⁰ (Another 117 did not meet their eligibility criteria to be reviewed and of the 39 reviewed,⁴¹ the designs of 30 were not up to snuff so their results were ignored though they might have played a role in the tangle!). What further assumptions make any or all of these relevant to your prediction? What makes effect sizes in these studies relevant to yours?

Jeremy Hardie and I have laid out two necessary and sufficient conditions for the same effect size to hold in your setting as in some others (Cartwright and Hardie, 2012):

- The mean value of the (net effect of) the support factors is the same in both.
- There’s a causal pathway from Read 180 to better comprehension scores in your setting.

The first depends on information about support factors which you are not given. The second, as we saw in Lecture 1, depends on the underlying structure in your school setting with your students. What evidence do you have that yours is sufficiently like that in the study setting, or alternatively, if it is different, that it is nevertheless the right kind to afford the Read 180 pathway? Also, which of the studies can you make a relevance argument for? They all had different results. Maybe though that is okay because all the results that were not indeterminate were positive. Is the fact that these 3 were positive relevant to the prediction that you will have positive results? If so, what are the additional assumptions that show this?

This would all be easier if we could just suppose what seems to be implied by ‘It works’— ‘Read 180 works (almost) everywhere’. It follows that it will work for you. But how good is the evidence for that? 3 studies? Or even many more? There are more than 2000 swans in the Thames, and in the right season, more than 40,000 in the UK. All white. Still in Sydney Harbour we know we can find black swans.

Alternatively, ‘Read 180 works’ is a loose precept (like ‘Democracies don’t go to war with democracies’), to be used as a tool in our toolbox. Now we are back to the problem I raised earlier: ‘[I]n the Read 180 case we have little know-how, no good, thick tangle of practices surrounding the development of predictions and explanations, whereas by now we do with the democratic peace.’ When I raise these issues about the thinness of the evidence base supplied in What Works sites for policy prediction, questioning even its claims to relevance, I am asked, ‘What’s the alternative?’ So I turned to thinking about why I was willing to trust the predictions that the Excimer Laser used in my eye surgery would emit a precise wavelength of

³⁹ <https://ies.ed.gov/ncee/wwc/intervention/742>. Accessed 2 April 2019.

⁴⁰ WWC Intervention Report on *Read 180*. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_read180_112916.pdf. Accessed 2 April 2019. Note that the 6 studies did include 3882 students in 61 schools.

⁴¹ Only 6 of the 9 reported results on comprehension.

cool ultraviolet light that would not burn my eye. It wasn't just the quantum theory that reassured me. It was the tangle of practices that I had studied in Stanford in theorising, designing and building lasers that gave me faith. But it won't help in the policy domain to use physics and engineering cases. 'Physics has true universal laws to work with. We don't.' That is why I have turned to the democratic peace. It is an exemplar of good social science theorising, with a thick tangle of practice that allows it to build good explanatory models of real conflict interactions between real countries. But how credible are these explanations?

Following my third element for underwriting credibility, my strategy would be to trust only models that are 'tried-and-true', that different practices converge on, that sit 'in the centre of the tangle.' This is familiar advice: trust most to what is overdetermined. And it is a core part of good social science methodology. We can turn to DPT for illustration. Here are some examples, all from Bruce Russett's short book *Grasping the Democratic Peace*:

- 'Even with the differing definitions of democracy and of war, this generalization is exceedingly robust.' [The generalization: in the 20th century 'war between democracies becomes impossible or almost impossible to find'.] (Russett, 1994, p. 20)
- Two sets of data [ICB and MID] on conflict don't treat all the same dyads. 'This does not mean that either set is inaccurate; rather, they vary enough in how they identify cases for us to use the two data sets as a check on how robust our results are.' (Russett, 1994, p. 76)
- 'As a further check we created an alternative scheme of regime categories from data of Arthur Banks...' (Russett, 1994, p. 77)
- The tables used usually presented data on the independent variable in continuous or categorical form. 'Most analysis was also done by analysing all the independent variables as dichotomous.' (Russett, 1994, p. 84)
- '[T]o be sure the outlier was not unduly responsible for the explanatory power of the political variables we re-estimated the equation without it.' (Russett, 1994, p. 107)

With this advice in mind what applications seem sound? As Russett reports, the US has used covert forcible actions against a number of elected governments of less-developed countries, for example Iran (1953), Guatemala (1954), Indonesia (1957), Brazil (1961), Chile (1973), Nicaragua (1981). Should we have bet against these on grounds of the Democratic peace? If we did, we should have known to hedge our bets since these were not wars under many of the characterisations of wars on offer in DPT, and 'these governments were not fully democratic according to the criteria that have been applied here...' (Russett, 1994, p. 121). Predictions against these covert actions do not come out of the centre of the tangle.

IN CONCLUSION

Middle-range laws, Elster-type mechanisms, maxims and precepts are fine in science. They are more than fine, they are a boon. We do not need to recast them as propositions, we do not need to take them literally, we do not need to try to 'derive' our conclusions from them. We do need to learn how to use them. We do need, as a community, to fill our toolbox with a big variety of tools and to learn how to use them together to craft reliable finished products.

I have set down three elements that make a claim to reliability credible. I expressed them in a loose, generic, metaphorical way that provides little guidance about what they look like in any particular case. I do not think that is a problem. Evaluating the credibility of a model is not an exact science. We may long for probabilities — they sure would help with cost-benefit analysis which is why the EBP community is keen on them. But we need a chance set-up, like a roulette wheel, to generate probabilities and there are no chance set-ups in the offing.

Of course this is a highly debated issue which I don't propose to take on in my concluding paragraphs. What I shall do instead in is to remind you of Neurath's boat, which is apt to my topic. Mauricio Suárez (2019) notes that the focus on practice is a return to the good old Vienna Circle Philosophy of Neurath, Hahn, Zilsel and Carnap. We are always building —not rebuilding— our ship at sea without ever having been on solid ground to start with. Neurath's point was that we have always to work from where, and with whom, we are. We must use the materials we have (Suárez, 2019), the methods and tools we have and the practitioners we have. These will generally disagree. That is the human condition. There is only the very practices and the very practitioners we have. There is no scientific method writ in some Platonic heaven that could show us the way to true claims. There is no magic decision procedure we can follow to arrive at best actions. For better or worse, there is just us and we must muddle it out.

Acknowledgements

Special thanks to Ann C Thresher, who has worked with me both on middle-range theory and on the democratic peace, as well as to the PhD students in the 2019 winter term seminar in philosophy of social science at UCSD, who read and commented very helpfully on these lectures, and to Haggeo Cadenas and Eugene Chua for editorial help. Much of the research was funded through the project Knowledge for Use (K4U), for which I am very grateful. K4U has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 667526 K4U). The content reflects only the author's view and the ERC is not responsible for any use that may be made of the information it contains.

Lullius Acknowledgements

Special thanks to Ann C Thresher, who has worked with me both on middle-range theory and on the democratic peace, as well as to the PhD students in the 2019 winter term seminar in philosophy of social science at UCSD, who read and commented very helpfully on these lectures, and to Haggeo Cadenas and Eugene Chua for editorial help. Much of the research was funded through the project Knowledge for Use (K4U), for which I am very grateful. K4U has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 667526 K4U). The content reflects only the author's view and the ERC is not responsible for any use that may be made of the information it contains.

REFERENCES

- Babst, D.V. (1964). Elective governments: A force for peace. *The Wisconsin Sociologist*, 3(1), 9-14.
- Baviskar, S., & Malone, M.F.T. (2004). What democracy means to citizens—and why it matters. *Revista Europea de Estudios Latinoamericanos y del Caribe/European Review of Latin American and Caribbean Studies*, 3-23.
- Beach, D., & Pedersen, R.B. (2019). *Process-tracing methods: Foundations and guidelines*. University of Michigan Press.
- Bradburn, N.M., Cartwright, N., & Fuller, J. (2017). A Theory of Measurement. *Measurement in Medicine: Philosophical Essays on Assessment and Evaluation*, 73-88.
- Bunge, M. (1997). Mechanism and explanation. *Philosophy of the social sciences*, 27(4), 410-465.
- Bunge, M. (2004). How does it work? The search for explanatory mechanisms. *Philosophy of the social sciences*, 34(2), 182-210.
- Cartwright, N.D. (2013). *Evidence: for policy and wheresoever rigor is a must*. London: London School of Economics.
- Cartwright, N. (2019). *Nature, the artful modeler: Lectures on laws, science, how nature arranges the world and how we can arrange it better*. Chicago: Open Court Publishing.
- Cartwright, N., Hardie, J., Montuschi, E., Soleiman, M., & Thresher, A. (forthcoming). *The Tangle of Science: Reliability without Rigour*. Oxford: Oxford University Press.
- Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford: Oxford University Press.
- Cartwright, N., Cat, J., Uebel, T.E., & Fleck, L. (2008). *Otto Neurath: Philosophy between science and politics* (Vol. 38). Cambridge: Cambridge University Press.
- Cartwright, N., Pemberton, J., & Wieten, S. (2018). Mechanisms, ceteris paribus laws and covering-law explanation. *Centre for Philosophy of Natural and Social Science Working Paper*. London: LSE.
- Cartwright, N., Shomar, T., & Suárez, M. (1995). The tool box of science: Tools for the building of models with a superconductivity example. *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 44, 137-149.
- Chang, H. (2017). Pragmatist coherence as the source of truth and reality. In *Proceedings of the Aristotelian Society* (Vol. 117, No. 2, pp. 103-122).
- Choi, S.W. (2016). A menace to the democratic peace? Dyadic and systemic difference. *International Studies Quarterly*, 60(3), 573-577.
- Cohen, R. (1994). Pacific unions: a reappraisal of the theory that 'democracies do not go to war with each other'. *Review of International Studies*, 20(3), 207-223.
- Crasnow, S. (2012). The role of case study research in political science: Evidence for causal claims. *Philosophy of Science*, 79(5), 655-666.
- Craver, C.F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford University Press.
- Davey, C., Hargreaves, J., Hassan, S., Cartwright, N., Humphreys, M., Masset, E. & Bonell, C. (2018). Designing evaluations to provide evidence to inform action in new settings. *CEDIL Inception Paper*, (2).
- Deaton, A. (2010). Understanding the mechanisms of economic development. *Journal of Economic Perspectives*, 24(3), 3-16.
- Di Bucchianico, M.E. (2009). *Modelling high temperature superconductivity: A philosophical inquiry in theory, experiment and dissent* (Doctoral dissertation, London School of Economics and Political Science (United Kingdom)).
- Dowe, P. (2000). *Physical causation*. Cambridge: Cambridge University Press.
- Efstathiou, S. (2012). How ordinary race concepts get to be usable in biomedical science: An account of founded race concepts. *Philosophy of science*, 79(5), 701-713.
- Eisenhart, M., & Towne, L. (2003). Contestation and change in national policy on "scientifically based" education research. *Educational researcher*, 32(7), 31-38.

- Elster, J. (2015). *Explaining social behavior: More nuts and bolts for the social sciences*. Cambridge: Cambridge University Press.
- Farber, H.S., & Gowa, J. (1997). Common interests or common polities? Reinterpreting the democratic peace. *The Journal of Politics*, 59(2), 393-417.
- Farnham, B. (2003). The theory of democratic peace and threat perception. *International studies quarterly*, 47(3), 395-415.
- Geva, N., & Hanson, D.C. (1999). Cultural similarity, foreign policy actions, and regime perception: An experimental study of international cues and democratic peace. *Political Psychology*, 20(4), 803-827.
- Harré, R. (1972). *The Philosophies of Science*. Oxford: Oxford University Press.
- Haugeland, J. (2004). Two dogmas of rationalism. In Adams, Z. (2017). *Giving a damn: Essays in dialogue with John Haugeland*. Cambridge, MA: MIT Press.
- Hedstrom, P., & Swedberg, R. (1998). *Social mechanisms*. Cambridge: Cambridge University Press.
- Hedström, P., & Swedberg, R. (1998). Social mechanisms: An introductory essay. *Social mechanisms: An analytical approach to social theory*, 1-31.
- Hedström, P., Bearman, P., & Bearman, P.S. (eds.) (2009). *The Oxford handbook of analytical sociology*. Oxford: Oxford University Press.
- Hegre, H., Bernhard, M., & Teorell, J. (2018). Reassessing the Democratic Peace: A Novel Test Based on the Varieties of Democracy Data. *V-Dem Working Paper*, 64.
- Hempel, C.G., & Neurath, O. (1952). *Fundamentals of concept formation in empirical science* (Vol. 2, No. 7). Chicago: University of Chicago Press.
- Hitchcock, C., & Sober, E. (2004). Prediction versus accommodation and the risk of overfitting. *The British journal for the philosophy of science*, 55(1), 1-34.
- Kant, I. (1917). *Perpetual peace: A philosophical essay* (M.C. Smith, Trans.). London: G. Allen & Unwin Ltd. (Originally published in 1795).
- King, G., Keohane, R.O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton: Princeton University Press.
- Leonelli, S. (2019). What distinguishes data from models? *European Journal for Philosophy of Science*, 9(2), 22.
- Longino, H.E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton: Princeton University Press.
- Mackie, J.L. (1965). Causes and conditions. *American Philosophical Quarterly*, 2(4), 245-264.
- Mahoney, J. (2001). Beyond correlational analysis: Recent innovations in theory and method. In *Sociological forum* (pp. 575-593). Eastern Sociological Society.
- Morgan, M. S. (2007). *The curious case of the prisoner's dilemma: model situation? Exemplary narrative?* (pp. 157-185). Durham, NC: Duke University Press.
- Munro, E. (2011). *The Munro review of child protection: final report, a child-centred system* (Vol. 8062). The Stationery Office.
- Owen IV, J.M. (2005). Iraq and the Democratic Peace-Who Says Democracies Don't Fight. *Foreign Affairs*, 84, 122.
- Owen, J.M. (1994). How liberalism produces democratic peace. *International Security*, 19(2), 87-125.
- Pawson, R. (2000). Middle-range realism. *European Journal of Sociology/Archives Européennes de Sociologie*, 41(2), 283-325.
- Pawson, R., & Tilley, N. (1997). An introduction to scientific realist evaluation. In Chelmsky, E., & Shadish, W.R. (eds.) *Evaluation for the 21st century: A handbook*, 405-418. Washington: Sage Publications, Inc.
- Pawson, R., Tilley, N., & Tilley, N. (1997). *Realistic evaluation*. Washington: Sage Publications, Inc.
- Pennock, J.R. (2015). *Democratic political theory*. Princeton: Princeton University Press. (Originally published in 1979).
- Ray, J.L. (1993). Wars between democracies: rare, or nonexistent?. *International Interactions*, 18(3), 251-276.

- Robert, D. (1971). *Polyarchy: participation and opposition*. New Haven: Yale University Press.
- Rosato, S. (2003). The flawed logic of democratic peace theory. *American political science review*, 97(4), 585-602.
- Russett, B. (1994). *Grasping the democratic peace: Principles for a post-Cold War world*. Princeton: Princeton University Press.
- Salmon, W.C. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
- Strevens, M. (2012). Ceteris Paribus Hedges: Causal Voodoo that Works. *Journal of Philosophy* 109, 652-675.
- Suárez, M. (2019). *Filosofía de la ciencia: historia y práctica*. Madrid: Tecnos.
- Tomz, M.R., & Weeks, J.L. (2013). Public opinion and the democratic peace. *American political science review*, 107(4), 849-865.
- Turnell, A. (2012). The Signs of Safety. *Comprehensive Briefing Paper*. Resolutions Consultancy.
- Vanhanen, T. (2007). *Vanhanen's Index of Democracy*. International Peace Research Institute, Oslo (PRIO).
- Weisiger, A., & Gartzke, E. (2016). Debating the Democratic Peace in the International System. *International Studies Quarterly*, 60(3), 578-585.
- Whelan, F.G. (1983). Prologue: Democratic theory and the boundary problem. *Nomos*, 25, 13-47.
- Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

NANCY CARTWRIGHT is Distinguished Professor, UCSD (In residence Winter term) and Professor of Philosophy, Durham University (In residence Michaelmas and Summer terms) and Co-Director of CHESS, Durham University.

ADDRESS: Department of Philosophy, Durham University, 50 Old Elvet, DH1 3HN, United Kingdom.
E-mail: nancy.cartwright@durham.ac.uk / ORCID: 0000-0002-0873-6966